



Quantum Markov Chain Mixing and Dissipative Engineering

Kastoryano, Michael James

Publication date:
2012

Document version
Early version, also known as pre-print

Citation for published version (APA):
Kastoryano, M. J. (2012). *Quantum Markov Chain Mixing and Dissipative Engineering*.



Quantum Markov Chain Mixing and Dissipative Engineering

PhD thesis
Michael James Kastoryano

The PhD School of Science
Quantum Information Group
QUANTOP – Danish National Research
Foundation Centre for Quantum Optics
Niels Bohr Institute
and Niels Bohr International Academy

Academic supervisors: Michael M. Wolf and Anders S. Sørensen

December, 2011

Quantum Markov Chain Mixing and Dissipative Engineering

Michael James Kastoryano

This thesis is submitted in partial fulfilment of the requirements
for the Ph.D. degree at the University of Copenhagen.

Copyright © 2011 Michael James Kastoryano

Abstract

This thesis is the fruit of investigations on the extension of ideas of Markov chain mixing to the quantum setting, and its application to problems of dissipative engineering. A Markov chain describes a statistical process where the probability of future events depends only on the state of the system at the present point in time, but not on the history of events. Very many important processes in nature are of this type, therefore a good understanding of their behavior has turned out to be very fruitful for science. Markov chains always have a non-empty set of limiting distributions (stationary states). The aim of Markov chain mixing is to obtain (upper and/or lower) bounds on the number of steps it takes for the Markov chain to reach a stationary state. The natural quantum extensions of these notions are density matrices and quantum channels. We set out to develop a general mathematical framework for studying quantum Markov chain mixing.

We introduce two new distance measures into the quantum setting; the quantum χ^2 -divergence and Hilbert's projective metric. The quantum χ^2 divergence allows us to extend the class of functional techniques called Poincaré inequalities to the quantum setting. In the process, we identify the appropriate framework for discussing notions such as detailed balance and χ^2 -mixing: monotone Riemannian metrics on matrix spaces. This insight allows us to characterize the connection between spectral properties of a (primitive) quantum channel and its mixing time. Within the same framework we also derive a restricted quantum version of the celebrated conductance bound (Cheeger's inequality). We then consider Hilbert's projective metric - a well known tool in Perron-Frobenius theory - in the context of quantum information theory. As a classic tool for analyzing convergence of positive maps on cones, Hilbert's projective metric is especially good at providing existence proofs of maps between two spaces (or cones). We relate this measure to distinguishability measures on restricted cones of operators, and analyze the observable loss of information after the application of a quantum channel under a restricted set of measurements. Various notions of contractivity of quantum channels are revealed through Hilbert's metric.

Still on the topic of quantum Markov chains, we introduce the notion of cutoff phenomenon to the quantum setting. The cutoff phenomenon describes the situation when a Markov chain does not converge for a potentially long time, and then at a specific point in time abruptly converges to equilibrium. In the thermodynamic limit, the convergence profile will then look like a step function. We show that this type of convergence behavior occurs, as measured in trace-norm, when a system is subject to a product channel, and the noise on each product is modeled by a one-parameter semigroup of quantum channels.

Finally, we consider three independent tasks of dissipative engineering. The first task, which is aimed at experimental realization in the near future, consists in dissipatively

preparing a maximally entangled state of two atoms trapped in an optical cavity. We show that this is indeed possible with very high fidelity, and in a short amount of time, using present day technology. We also show that the scaling of the fidelity with the quality factor of the cavity (the cooperativity) scales quadratically better in our dissipative setup than in any known coherent unitary setup; indicating that dissipative state preparation can lead to fundamental improvements over closed system protocols. The second task that we consider is dissipative preparation of graph states, where we show that this class of states can be prepared in a time scaling as $\log n$ in the number of stabilizer elements. We also show that this process exhibits a cutoff. As a third task of dissipative engineering, we revisit the dissipative quantum computing construction of [VWC09]; we rigorously prove that it is as efficient as the circuit model.

Dansk resumé

Denne Ph.D.-afhandling omhandler videreudvikling af ideer angående konvergens af Markov-kæder i kvantemekanikken samt dens anvendelse inden for dissipativ manipulering. Markov-kæder beskriver en diskret tidslig proces for hvilken sandsynligheden af fremtidige begivenheder kun afhænger af systemets nuværende tilstand og ikke dens historik. Mange centrale processer i naturen er af denne type, hvorved en god forståelse af Markov-kædernes opførsel, har haft stor betydning for videnskaben. Markov-kæder har altid et ikke tomt sæt af stationære tilstande. Formålet med studier af Markov-kæders konvergens er at fastslå øvre og nedre grænser for antallet af skridt det tager for Markov-kæden at nå en stationær tilstand. Den naturlige kvantemekaniske udvidelse af disse begreber er tæthedsmatricer samt kvantekanaler. I denne afhandling ønsker vi at udvikle en general ramme for analyser af Markov-kæders konvergens.

To nye afstandsmål bliver introducerede, den kvantemekaniske χ^2 -divergens samt Hilberts projektive metrik. Den kvantemekaniske χ^2 -divergens muliggøre udvidelsen af funktionelle metoder kaldet Poincaré uligheder til kvantemekanikken. I den forbindelse identificere vi de nødvendige værktøjer for at diskutere begreber som detaljerede balance eller χ^2 -blanding: monotone Riemannske metrikker på matrix run. Dette muliggøre en karakteristik af forbindelsen imellem de spektrale egenskaber af en primitiv kvantekanal og dens konvergenstid. Inden for samme ramme udledes en begrænset version af den berømte grænse for ledningsevnen, Cheeger's ulighed. Efterfølgende studeres Hilberts projektive metrik - der er et velkendt koncept i Perron-Frobenius teori - i sammenhæng med kvanteinformationsteori. Som et af de klassiske værktøjer til at analysere konvergens af positive funktioner på kegler, er Hilberts projektive metrik i særdeleshed god i forbindelse med eksistens beviser for afbildninger imellem to rum (eller kegler). Dette afstandsmål relateres til forskellighedsmål på begrænsede kegler af operatorer, og analyser af observablens tab af information efter anvendelse af en kvantekanal under et begrænset sæt af målinger. Forskellige forståelser af sammentrækningen af kvantekanaler afsløres igennem Hilberts metrik

Stadig i forbindelse med kvantemekaniske Markov-kæder introducere vi begreber af afbrydelsesfænomener i rammerne af kvanteteorien. Afbrydelses fænomener beskriver situationen hvor en Markov-kæde ikke konvergere for en potentiel lang tid, hvorefter den for en specifik tid brat vil konvergere mod sin stationære tilstand. I den termodynamiske grænse, vil konvergens profilen ligne en trappe funktion. Det vises at denne type af konvergens opstår, målt af spor normen, når systemet udsættes for en kvanteproduktkanal samt at støjen til ethvert punkt kan modelleres som en en-parameter semigruppe af kvantekanaler.

Til sidst beskæftiger vi os med tre uafhængige opgaver i forbindelse med dissipativ manipulering. Først opgave, med henblik på snarlig eksperimentel virkeliggørelse, består

i en dissipativ forberedelse af en maksimalt sammenfilteret tilstand imellem to atomer fanget i en optisk cavity. Vi viser at det med nutidens teknologi er muligt at frembringe en sådan tilstand på kort tid og med høj nøjagtighed. Det vises også at nøjagtigheden skalere kvadratisk bedre i det dissipative tilfælde sammenlignet med kendte kohærente unitære systemer. Dette er en indikation på at dissipativ tilstands forberedelse kan lede til fundamentale forbedringen i forhold til lukkede systemer.

Den anden opgave, som vi antager, er dissipativ forberedelse af graftilstander, her vises det at disse tilstand kan forberedes med en tids skalering som $\log n$ i antallet af stabilisator elementer. Det vises også at denne processer har en afgrænsning. For den tredje opgave i dissipativ manipulering genovervejer vi den dissipative kvante computer af [VWC09]; det bevises stringent at denne model er ligeså effektiv som den almindelige kvantecomputere.

Acknowledgments

This thesis concludes three years of Ph.D. studies carried out jointly at the Niels Bohr Institute, University of Copenhagen, and at the Niels Bohr International Academy. I gratefully acknowledge the generous support by the faculty of the Niels Bohr Institute and the Niels Bohr International academy, and of the European projects QUEVADIS and COQUIT.

These have been three wonderful years in a highly stimulating, as well as gratifying, environment. Naturally, the personal and professional educational process which culminates with this thesis would not have been possible without the help and support of a number of people to whom I will for a long time feel indebted.

My first thanks go to my principal supervisor, Michael M. Wolf who took me under his wing in a period of my life filled with doubt and confusion. He guided me through the misty beginnings of a PhD project, gave me the trust that I needed to materialize my first projects, and taught me how to think for myself. I am also very thankful to my co-advisor Anders Sørensen who acted as a perfect counterbalance, and took over the responsibilities when Michael Wolf went on parental leave, and then again when the group moved to the Technical University in Munich. With opposite styles and very different, but complementary, ways of thinking, each of my supervisors constitute wonderful role models, whose positive impact will stay with me for a long time to come. Special thanks also goes to my third mentor, Andrew Jackson, who contributed just as much to making me a physicist, even though we never collaborated on a project together. Through many delightful tea-break discussions, Andrew conveyed to me the importance and delights of thinking about a physical problem in a free and honest manner. This lesson is invaluable.

Furthermore, I am grateful for the numerous fruitful collaborations that have emerged with so many interesting people during my time as a graduate student. I would like to thank everybody I have had a chance to collaborate with and the pleasure to learn from in the past three years. In particular, David Reeb and Florentin Reiter, each of which has co-authored two papers with me, and have been great scientific companions. Frank Verstraete and Jens Eisert and their groups who warmly welcomed me as a guest on several occasions. Finally, Kristan Temme, and more recently Fernando Pastawski, whose scientific interests most closely match mine, and who have also been great company, both scientific and social, during the many mutual visits and encounters at conferences.

It is also a pleasure to thank all of the people who have populated the landscape of my professional life in the past three years. All of the members of the eclectic Copenhagen-Wolf group through time: Teiko Heinosaari, Viktor Eisler, Savanna Sterling, Max Schlosshauer, Anastasia Jivulescu, Alejo Salles and David Reeb. As well as those from the Sørensen group: Martijn Wuubs, Dirk Wittaut, Jonathan Bohr Brask, Eran Kot, Florentin Reiter

and Anna Grodecka-Grad. The warm and dynamic QUANTOP crew, who have made my workplace a fun and altogether delightful place to be, as well as the many academy members who helped me get out of the Quantum Optics/Quantum Information cocoon at times.

And last but not least, I thank my family and friends whose support, faith and encouragement has kept me afloat throughout the eternal challenge of growing up!

List of publications

The main results presented in this thesis have been published in the following papers:

- K. Temme , M. J. Kastoryano, M. B. Ruskai , M. M. Wolf , and F. Verstraete, “The χ^2 -divergence and Mixing times of quantum Markov processes”, *J. Math. Phys.* **51**, 122201 (2010).
- M. J. Kastoryano, F. Reiter, A. S. Sørensen, “Dissipative preparation of entanglement in optical cavities”, *Phys. Rev. Lett.* **106**, 090502 (2011).
- D. Reeb, M. J. Kastoryano, and M. M. Wolf, “Hilbert’s projective metric in quantum information theory”, *J. Math. Phys.* **52**, 082201 (2011).
- M. J. Kastoryano, D. Reeb, M. M. Wolf, “A Cutoff Phenomenon for Quantum Markov Chains”, arXiv:1111.2123 (2011) (Submitted to Journal of Physics A).

Other publications by the author which are not included in this thesis:

- C. Degueldre, M. J. Kastoryano, and K. Dardenne, “Variable incidence angle X-ray absorption spectroscopy for the study of Zircaloy corrosion layers”, *J. Nucl. Mater.* **362** 316-326 (2007).
- M. Kliesch, T. Barthel, C. Gogolin, M. Kastoryano, and J. Eisert, “A dissipative quantum Church-Turing theorem”, *Phys. Rev. Lett.* **107**, 120501 (2011).
- F. Reiter, M. J. Kastoryano, A. Sørensen, “Entangled steady-states of two atoms in an optical cavity by engineered decay”, arXiv:1110.1024 (2011) (Submitted to The New Journal of Physics).

Contents

Abstract	iii
Dansk resumé	v
Acknowledgments	vii
List of publications	ix
1 Introduction	3
1.1 Chapter summaries	12
2 Fundamentals	15
2.1 Basic notions	15
2.1.1 Norms	15
2.1.2 From classical to quantum probability: quantum Markov chains .	17
2.2 Spectral properties of channels	18
2.2.1 Fixed points of quantum channels	21
2.3 Distance measures	22
2.3.1 Distance measures review	22
2.4 Contraction and convergence measures	25
2.5 The Convergence Theorem	27
3 Quantum Markov chain mixing	31
3.1 Markov chain mixing	32
3.2 The quantum χ^2 -divergence	32
3.2.1 Monotone Riemannian metrics and generalized relative entropies.	34
3.2.2 Properties of the quantum χ^2 -divergence	36
3.3 Mixing time bounds and contraction of the χ^2 -divergence.	38
3.3.1 Mixing time Bounds	38
3.3.2 Contraction Coefficients	41
3.4 Quantum Detailed Balance	44
3.5 Quantum Cheeger's Inequality	47
3.5.1 Example: Conductance bound for unital qubit channels	49
4 The Cutoff Phenomenon	51
4.1 The Cutoff Phenomenon	52
4.2 Main Results	53
4.3 Examples of the cutoff phenomenon and applications	56
4.3.1 Primitive Channels with Separable Initial States	57

4.3.2	Channels with Unique Pure State Fixed Point	57
4.3.3	Qubit Amplitude Damping	59
5	Dissipative Engineering	61
5.1	Dissipative state preparation	61
5.1.1	Dissipative Preparation of Graph States	67
5.2	Dissipative computation	68
5.2.1	Dissipative quantum computation on $\mathcal{H}_2^{\otimes N} \otimes \mathcal{H}_{M+1}$ with quasi-local Lindblad operators	68
5.2.2	Unary encoding of the time register, i.e DQC on $\mathcal{H}_2^{\otimes(N+M)}$	72
6	Hilbert's projective metric	77
7	Summary and outlook	111
7.1	Summaries	111
7.2	Outlook	113
	Bibliography	115

For Katherine, my Mom, to whom I owe my curiosity.

Introduction

Quantum information science is now a well established field of physics which has seen some remarkable successes since its inception a decade and a half ago. This progress has been fueled from a number of different directions including pure mathematics, theoretical and experimental physics, computer science, engineering and philosophy. Its, sometimes overdone, promises of super-computers of the future, along with proposals of teleportation and perfectly secure cryptography, have fueled the imagination of science-amateurs, and the generosity of funding agencies. I would argue, though, that this enthusiasm has not been artificially inflated, and hat looking back at 15 years of research, a number of fantastic results and deep insights have been gained. It is in this fruitful young field that this thesis finds its foundation.

We go through a brief geographic tour of Quantum Information Science, discussing some of the main successes so far, and some of the major challenges for the near future. Quantum Information Science has its origins already in the basic debates around the philosophical implications of quantum mechanics, which culminated in the famous 1935 Einstein-Podolski-Rosen [EPR35] paper and the responses to it by Bohr [Boh35] and Schrödinger [SB08], which alluded to the notions of entanglement and hidden variable theories. A decade or so later, Von Neumann, Turing and a number of others [Tur37, Von45] established theoretical computer science, and a few years after that Shannon founded information theory *ex nihilo* [Sha01]. Hence, by 1950 the three cornerstones of theoretical quantum information science (quantum theory, computer science and information theory), were already in place. It would however take another 40 years before people really took the possibility of processing quantum information seriously. The reason is twofold. On the one hand, after the generation which founded quantum mechanics, physicists became increasingly pragmatic about the theory, considering it a wonderful tool for calculations, but not reading into its meaning (and hence potential) much further¹.

On the other hand, experiments had not reached a point of control and precision that would require anything more than the standard quantum theory to describe. This changed however around 1990, when a series of experiments at the frontier between quantum optics

¹Of course I am not quite doing 60 years of physics justice, and the wonderful accomplishments, such as the Standard Model, or superconductivity could not have been reached without a deep understanding of quantum theory, but the now popular adage of "shut-up and calculate"[Mer04] has to a large extent prevailed as the *modus operandi* of the practicing physicist.

and condensed matter physics started being able to isolate and manipulate single quantum systems (ex: trapped ions [LBMW03, WBB⁺92], Cavity QED [DRBH95, THL⁺95],...), and coherent mesoscopic systems (ex: BEC [CW98, AEM⁺95, DMA⁺95], quantum dots [WBM⁺95, LCW⁺96],...). With these technological advancements, and the final acknowledgement (as a result of the brilliant insight of John Bell [Bek66] and its masterful implementation by Aspect et al. [AGR81, ADR82, AGR82]), that non-locality and entanglement were an intrinsic facet of physical reality, a new quantum theory was not only possible, it was necessary. By new quantum theory, I do not mean an extension, in the sense that supersymmetry hopes to extend gauge field theory, or quantum gravity as an extension of general relativity. The new quantum theory, which people now just call *quantum information*, is an emergent theory, in the same way as synthesized genetic engineering is a scientific field emerging from the theory of DNA structure and technological maturity.

In the early 1990's, a few theoretical proposals [CZ95, LD98] and early experiments [MMK⁺95, DRBH95] clearly started indicating that Feynman's late claim [Fey82] of a quantum machine to solve quantum problems might not be all that wild an idea. The field of Quantum Information Science [NC00], however, really took off in 1994 when Peter Shor proposed a quantum algorithm for integer factorization [Sho94] and proved that it was efficient (no efficient classical algorithm exists today). Ever since Shor's discovery, the field has been growing and sprawling into ever distant realms of the physics landscape. Rather than trying to go through all of the areas where quantum information has left its fingerprint, we illustrate in Figure 1.1 a geographic layout of the present state of our science.

The figure is not meant to be taken too seriously, but it should illustrate some of the key features of Quantum Information Science (QIS) which perhaps distinguish it from the other fields of physics. As mentioned above, the field is fundamentally interdisciplinary as a result of its intrinsic structure. Indeed, whereas other fields of physics are defined either by a common set of methods or a type or scale of system, quantum information in many ways stems from the dream of one day seeing a scalable quantum computer on our desktops. Subsequently, a number of basic principles have been synthesized as pertaining specifically to QIS², but as can be seen in the figure, no single concept or method can consistently string together the wide range of topics which fall under the umbrella of QIS. What also makes the field so peculiar is that we not only do not know how to build such a quantum machine, but we also do not really know what it could be good for (besides factoring large numbers). It is in some sense this tension between technological incentive and fundamental - almost philosophical - inquiry which has made QIS such a fertile landscape for new ideas.

It should be noted that almost all of the theoretical topics depicted in Figure 1.1 are based on a closed system description; a quantum algorithm is described as a succession of unitary gates, quantum states are often represented as ground or thermal states of Hamiltonians, etc. Of notable exception are the theories of *quantum error correction*, *fault-tolerance* and *decoherence* where the effect of the environment plays a crucial role. However, in all of these cases, the effects on the environment are studied in order to engineer systems

²The firm belief that information transmission and processing is independent of the physical medium which carries it, and can be studied abstractly (reminiscent of Shannon's notion of information), is one such basic principle.

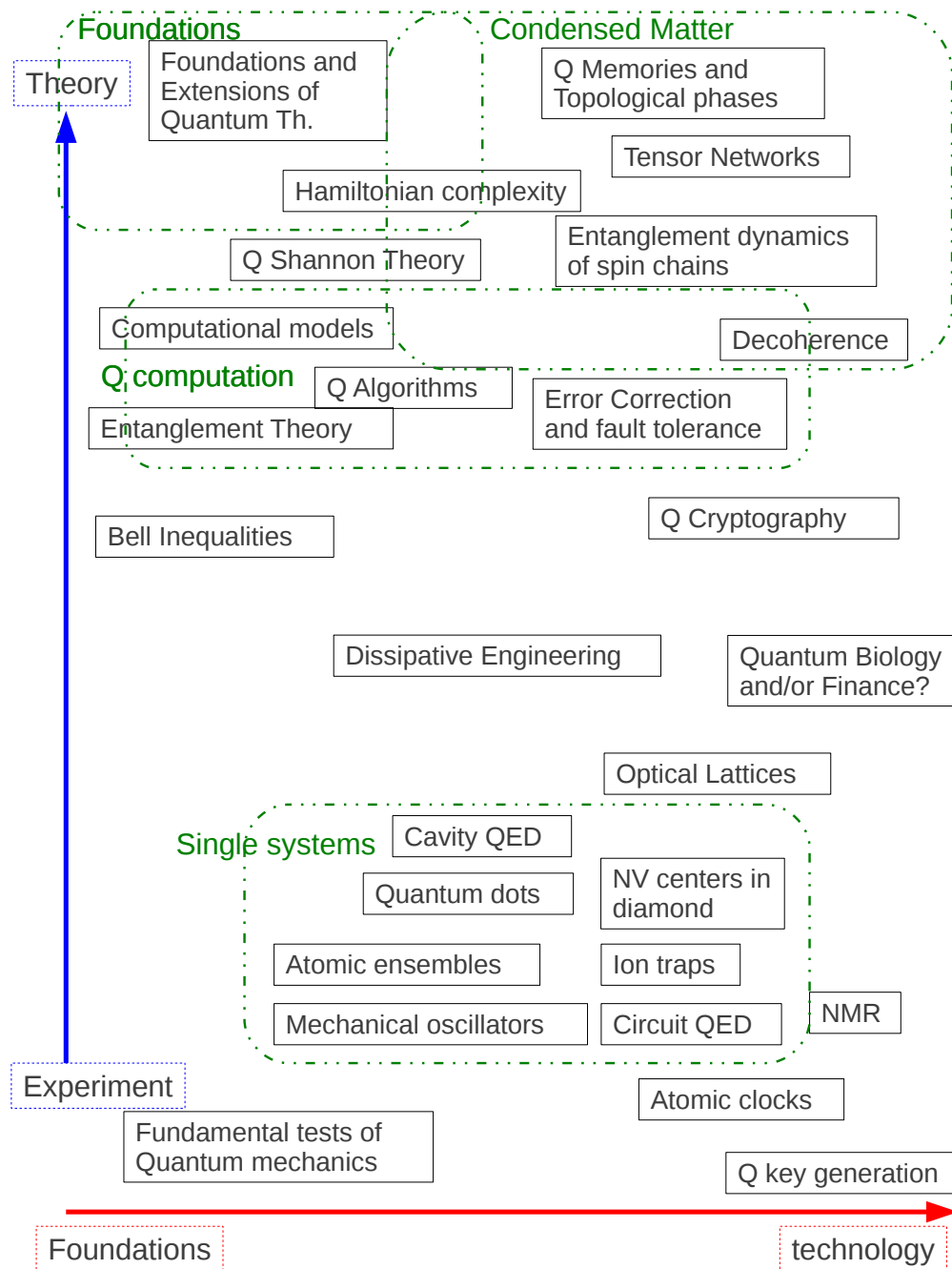


FIGURE 1.1: The above figure is meant to characterize the geography of quantum information science on an *Experimental to Theoretical* and on a *Foundations to Technology* scale. Each black box represents an active topic in quantum information science. The green dot-slashed boxes are meant to indicate roughly where the different topics stand with respect to general fields of study. What topics are closer to technological implementation could be source of heated debate, as no one can know beforehand what discoveries have fruitful applications. Clearly the choice and placement of topics is particular to the author's personal appreciation of the field and does not reflect strict common consensus within the community.

which are less vulnerable to noise. The only two labeled topics where the environment plays an active part in the quantum coherence are *Quantum Shannon Theory* and *Dissipative Engineering*. Our starting point is with the latter, where the idea is to develop protocols which exploit dissipative dynamics. Indeed, given that a quantum system inevitably interacts with its environment, and most often more than we would like, why not exploit the extra degrees of freedom offered by nature to process information?

One of the major bottlenecks in understanding the information processing potential of open quantum systems has been the lack of quantitative tools for analyzing their convergence to stationarity. The interest in understanding the convergence behavior of quantum processes has cropped up in several distinct areas in Quantum Information Theory and many body physics, and seems to require a new set of tools for its analysis. Some of these instances are: (i) Dissipative State Preparation, (ii) Dissipative Quantum Computation, (iii) Quantum Memories, (iv) Quantum Monte-Carlo type algorithms, and (v) correlation lengths in Finitely Correlated and Tensor network states. In all of these cases, and in many more, one would like to have quantitative tools for upper and/or lower bounding the time to stationarity.

We discuss in some more detail how convergence analysis of quantum channels enters in the picture. Some familiarity with notions of quantum information theory and computer sciences are assumed from this point on. These notions will be properly introduced in Chapter 2.

- **Dissipative State Preparation (DSP):** The main goal of DSP, as enunciated recently in [VWC09, KBD⁺08, DMK⁺08], is to engineer a physically realistic system described by a Markovian master equation, whose unique (pure) stationary state is the one to be prepared³. The Markovian master equation assumption is typically a very good one in quantum optics. We point out that the system's dynamics do not have to be purely dissipative, and in fact understanding the interplay between the dissipative and the unitary dynamics is a very important problem which has only been addressed very summarily.

There are two main approaches to DSP. The first, which could be called the condensed matter approach, is to map out the stationary-state phase diagram of some class of master equations, and study its many body properties as a function of an appropriate set of (physically motivated) free parameters in its Hamiltonian and Lindblad operators [TDZ11, DTM⁺10]. One might also be interested in the transient dynamics, which sometimes characterize more of the relevant physics, as is the case for spin glasses [FH91]. It turns out that there is a close, but not fully understood, connection between the critical regions of the phase diagram and slow convergence to stationarity. More specifically, there is a correspondence between correlation lengths of spatially separated observables in the stationary-state and the convergence speed of the dynamics; i.e. two point correlations decay exponentially fast in the separation length iff the mixing time is fast (polynomial in the size of the system). Hence, one can get access to the static properties of the stationary phase diagram by studying the dynamics of the system and visa versa, so that

³One can think of more general settings for DSP. For instance, allowing for non-Markovian channels, or for preparing an auxiliary target state from which the desired state can be efficiently prepared. We stick to principles formulated in [VWC09, KBD⁺08, DMK⁺08], as the material in this thesis does not extend beyond them.

studying mixing times of time-continuous quantum channels whose generator is a Markovian master equation becomes a problem of many body physics. In particular, characterizing and understanding the roles of locality and frustration for open system dynamics is an important task.

The second approach, which could be called the quantum information approach, is to consider a class of states - for instance Matrix Product States (MPS) [PGVWC07, VMC08] - with certain properties (translation invariance, bounded bond dimension, etc.) and build master equations which (approximately) have these desired states as their unique (or quasi-unique) stationary state [VWC09, KBD⁺08]. Several important classes of states, which are generally referred to as Tensor Network States (TNS) [Vid07, VC04, CV09], have been defined and studied recently, in an effort to characterize, on the one hand, the physically relevant states in Hilbert space (i.e. those which are ground states of local frustration free Hamiltonians), and on the other hand, the states which can be easily simulated on a classical computer. It turns out that DSP is an excellent, and in some cases the only known, procedure for preparing TNS states; this might not come as a very big surprise, as nature itself prepare states by "cooling". Here again there is a trade-off between how rapidly these states can be prepared by DSP and how much entanglement they can contain, so that studying mixing times can address issues of many body entanglement as well.

- **Dissipative Quantum Computation (DQC)** is a method of performing quantum computation by engineering the dissipation between a system its environment [VWC09]. The gates, as well as the input state, of the computation are encoded into the Lindblad operators of some engineered master equation in such a way that the (quasi-)unique stationary state encodes the outcome of the computation and is reached in a time which grows only polynomially in the system size. Hence, guaranteeing that computation is efficient becomes a question of bounding the mixing time.

DQC offers a new avenue for tackling unsolved problems in Quantum complexity theory [Wat08] (ex: Hamiltonian Complexity [Osbl1]), as it provides us a whole new framework to work with. In particular, within this framework we can much more freely consider quantum irreversible computation, as well as hybrids of reversible and irreversible computation. This is perhaps not very surprising as many important problems in classical computational complexity are studied using techniques from Markov chain mixing, of which quantum mixing is a generalization.

- Finding good, let alone realistic, **Quantum Memories** is a formidable task, and one of the most important open problems in quantum information and computation theory [AHHH08, AFH09, CLBT10, BH11]. What one requires from a *good* quantum memory is the ability to store (encode) arbitrary quantum information for a *long* time with not too large an amount of classical information processing and local measurements. Ideally one would like not to have to perform any recovery operations at all and hope that the encoding has some form of topological protection against noise below some threshold. No such physically realistic encoding has been found yet!

Tools for studying mixing time could allow an attack on this problem from two different directions. One direction is addressing the question of whether a given

Hamiltonian is topologically stable against thermal noise by rephrasing it as a question about the mixing time of the Davies map (the quantum analogue of Glauber dynamics). This way one could obtain no-go theorems for certain classes of Hamiltonians. In particular, the relationship between the energy barrier and the mixing time is not well understood. The other direction is to merge the encoding, measurement and classical information processing together and exercise dissipative correction of quantum information. Preliminary results have been obtained in [PCC11], and classical analogous results are promising in this direction [Gac86, Gac01].

- Yet another area where mixing times play a crucial role is in **quantum Monte-Carlo sampling** and related algorithms. The goal of quantum Monte-Carlo type algorithms is to sample from the Gibbs state of some local quantum Hamiltonian. Recently, several such algorithms have been proposed [TOV⁺11, PW09, TD00, RGE11], and they seem to fall into two distinct categories. On the one hand there are those which have a runtime which is certifiable, but where for relevant problems the runtime is certifiably long [PW09, RGE11]. And on the other hand, there are the algorithms where the runtime is suspected to be fast, but where there is no way of certifying it [TOV⁺11, TD00]. Mixing time tools were developed in the classical context largely for answering such questions, and it is likely that the same intimate connection holds in the quantum case.
- Finally, a not so obvious application of quantum mixing time tools is in studying **correlation lengths of Finitely Correlated States** [FNW92]. Indeed, a subset of TNS can be described in a natural way in terms of a cpt map. The cpt map then contains all the information about the symmetries, the locality, and the entanglement in these states. Furthermore, there exists a theorem relating the mixing time of the cpt map describing the FCS to the correlation behavior of these states. Hence, exotic mixing time behavior could translate to exotic correlation behavior [PGVWC07, VMC08, VC04, CV09].

There are a number of other problems, such as open system dynamics of spin chains [SP10, Pro11], disorder in quantum systems [BO07, BEO09], continuous time information theory [KRW11], and many more which can benefit from tools from quantum mixing times. Naturally, the best way of developing new tools is in studying specific problems, so that the broader the problem set is, the more diversified the toolbox will be. **It is the aim of this thesis to develop the foundations of a theory of quantum Markov chain mixing, and look at a few examples of where it can be used.**

Finding tight mixing bounds on explicit physical processes is usually extremely difficult, and often times, the tools from (quantum) mixing times provide more information on the connection between the structure of a class of problems and its convergence behavior than on the mixing of a specific instance of the problem. A particularly striking example is the classical Ising model, where there exists a rigorous correspondence between rapid convergence in time under Glauber dynamics, and exponential decay of correlations in space.

Survey of classical mixing times:

As a first step in developing tools for studying convergence rates of quantum processes, it is important to understand how much of the machinery from the classical Markov chain

mixing literature can be borrowed. Indeed, by noting that a finite-dimensional quantum channel is the non-commutative analogue of the probability transition matrix of a finite-state Markov chain, many results from the field of Markov chain mixing can be translated to the quantum setting, with appropriate modifications.

Classical Markov chain mixing usually refers to a set of tools developed for bounding the convergence time of finite Markov chains [LPW08]. There have been extensions to countably infinite chains and some analogous continuous systems, but the bulk of the results pertain to finite discrete systems. This is the case mainly because of the problems considered: Markov chain Monte Carlo sampling, and combinatorial problems.

A few of the major results to date in classical mixing times are: (i) Finding a randomized algorithm for approximating the permanent of a matrix with positive entries, and *proving* that it converges rapidly [JS88, JSV04]. (ii) Finding a randomized algorithm for approximating the volume of a convex body, and *proving* that it converges rapidly [DFK91]. (iii) Characterizing the mechanisms of card shuffling (random walks on groups) [AD86, BD92] and proving that in almost all relevant cases, the convergence is rapid and exhibits a sharp *cutoff* at a specific point in time, when the system size becomes large [SC04]. The prototypical example is that "riffle" shuffling a deck of n cards which takes exactly $\frac{3}{2} \log n$ shuffles as n becomes large. (iv) Full analysis and characterization of the convergence behavior of the 2D Ising model with Glauber dynamics and various boundary conditions [Mar94, MO94, MOS94]. In particular, a number of fundamental results connecting mixing times and correlation lengths were obtained. In fact cutoff has even been shown in the rapidly mixing regime [LS09]. (v) The Propp-Wilson algorithm which allows to sample **exactly** from a distribution in polynomial time with only local addressing [PW96]! Some applications of this method have been considered, in particular for the dynamics of the Ising model, where qualitatively different behavior has been shown to occur as compared to Glauber dynamics [HN99]. The Propp-Wilson (or coupling from the past) method has also been successfully applied to the Hard-core model [PW97].

A plethora of different methods have been used to solve the above and many other problems which could be formulated in the language of Markov chain mixing. Below, we go through some of the main methods of Markov chain mixing, and briefly discuss possible quantum generalizations, or obstacles thereto. We also point to what progress has been made in this thesis in extending the classical results to the quantum domain.

- **The coupling method** is perhaps the most widely used method for bounding the mixing time of a Markov chain, as it offers a simple upper bound on the trace norm of the difference between two probability distributions, without any additional assumption about the chain or the fixed point. The basic idea behind the method is to consider two coupled Markov processes (i.e. that have the property that if at some point in time they both have the same value, then they will have the same value for every subsequent point in time), and bound the mixing time by the expected time for the two coupled chains to meet. This method can "provide spectacular results from pure thought" [Dia11], but does not allow for any systematic analysis.

More severely, the coupling method, and its generalizations to *path couplings* [Jer03] and *coupling from the past* [PW97] do not have any obvious generalization to the quantum setting. The reason for this is that they are based on probability arguments which have no apparent quantum counterpart.

- **The bottleneck ratio** (or conductance) is a technique which estimates whether the graph representing the probability transition of the Markov chain contains a bottleneck [DS91]. This method turns out to be quite good at obtaining upper bounds on very unconstrained (homogeneous) systems such as expander graphs, or at obtaining good lower bounds for systems which have an obvious bottleneck, such as coloring the star graph.

This technique has been partially generalized to the quantum setting, in the sense that the main theorem relating the bottleneck ratio holds for unital quantum channels [TKR⁺10]. We are unable to prove the theorem for general quantum channels because a critical step in the classical proof relies on the fact that there exists a preferred basis to work with (i.e. that the stationary state is diagonal in the physical basis), which is not always the case quantum mechanically.

- The χ^2 or **Poincaré type inequalities** are a set of functional techniques which are based on the connection between the χ^2 -divergence and the spectral gap of a reversible (detailed balance) stochastic matrix [Fil91]. This method in particular illustrates the close connection between spectral gap and convergence rates of Markov chains, and has been successfully applied to the exclusion process and to variants of it.

The basic framework of this method has been generalized to the quantum setting in [TKR⁺10], where we prove the analogous basic theorems, and show that the quantum case allows for a lot more freedom than the classical one. This freedom is best expressed within the framework of monotone Riemannian metrics on matrix spaces.

- Another set of functional tools, which can be seen as a strengthening of the Poincaré type inequalities, are the **Logarithmic Sobolev inequalities** [Dia96]. This technique uses the relative entropy as its basic distance measure, and hence accesses much more precisely the information content (or losses) in the dynamics of the processes. Most of the sharp classical mixing results with statistical physics models (Ising-type) are based on this technique [MO94, LS09].

As this is a functional technique, the theory can be formulated quantum mechanically. In addition, this technique is closely related to another well known tool called hypercontractivity, which quantifies the errors associated to coarse graining of noise acting on a system, and has been applied in the quantum setting in [CL93, RW96, MO08].

- In addition to the above tools for bounding the mixing time of specific Markov chains, a number of **comparison theorems** exist [DSC93, RT00], which allow one to bound the mixing time of a complicated chain with that of a simpler Markov chain whose mixing time is known.

In practice, for relevant problems which are not of combinatorial origin, the problem one wants to solve is always reduced to some easier problem which has a simple graph theoretic interpretation. Unfortunately, most of these comparison techniques rely heavily on properties of a specific graph, which makes quantum generalizations difficult, as there we want to make basis independent statements.

- Finally, a noteworthy topic is the **cutoff phenomenon**. The cutoff phenomenon is not a method for bounding convergence, but rather a type of convergence behavior exhibited by many Markov chains [Dia96]. It characterizes the situation when for some, possibly long, amount of time the Markov chain does not converge at all, and then suddenly, at a specific point in time it converges abruptly. This behavior has been proved to occur in many card shuffling examples.

This behavior has recently been seen to also occur in quantum mechanical examples consisting of product channels [KRW11].

It is worth noting that interweaving all of these techniques is a rich framework which allows for precise statements of theorems and conjectures. For a good general introduction to Markov chain mixing, see [LPW08]. For a very extensive exposition of mixing bounds and techniques, consult the lecture notes by Aldous and Fill [AF]. A good survey of the uses of mixing time tools to the problem of Markov chain Monte-Carlo sampling is [Dia98].

Previous work:

We briefly discuss some of the previous, and contemporary work, which has been done in the direction of quantum mixing times and dissipative engineering. Dissipative state preparation has become a popular topic recently, and in many ways was the motivation for much of the work in this thesis. The first set of papers, defined the problem and analyzed the preparation of Stabilizer states and Matrix Product States [KBD⁺08, VWC09] where the main focus was to convey the idea that dissipation can be helpful if catalyzed properly, and to show in principle how to prepare these classes of quantum states. In [VWC09], the authors defined dissipative quantum computation, and provided a sketch of the proof that it is at least as powerful as circuit quantum computation. A number of studies have come out proposing ways of preparing interesting, often entangle states as stationary states of dissipative systems in physically realizable setups. The systems of choice have been: optical lattices and Bose-Einstein condensates [DMK⁺08], Cavity QED and related setups [KRSr11, RKSr11, BDI⁺11, WS10], Atomic Ensembles [MPC11], trapped ions [BMS⁺11], Rydberg atoms [WML⁺10], NV centers in diamond [LGL⁺11]. The proposals of dissipative (entangled) state preparation in atomic ensembles and trapped ions have in fact been verified experimentally [LHN⁺11, KMJ⁺11]. Recently, an interesting topological state of matter was shown also to be preparable by dissipation [DRBZ11].

One of the exciting developments in dissipative engineering is that a number of quantum information tasks can be cast in the dissipative framework: dissipative quantum simulation [BMS⁺11], dissipative quantum repeaters and distillation [VMC11], dissipative quantum memories [PCC11].

On the other hand, very little has been done in the direction of generalizing Markov chain mixing tools to the quantum setting. A notable exception is the work of Olkiewicz and Zegarliniski on non-commutative Hypercontractivity and Log-Sobolev inequalities, and references therein [OZ99]. Finally, as alluded to earlier, a number of proposals have appeared exhibiting different ways of preparing Gibbs states on a quantum computer. These include quantum Metropolis sampling type algorithms [TOV⁺11], and various others [PW09, TD00, RGE11].

1.1 Chapter summaries

Chapter 2:

In this chapter, we set the notation, and we introduce some of the basic notions which will be used throughout the thesis. We define classical Markov chains, and propose quantum channels (cpt maps) as the natural quantum generalization of their probability transition matrix. We go on to describe the basic spectral properties of quantum channels. We define the set of rotating points of a channel and define the notions of irreducible and primitive channels. Next, we characterize the set of fixed points of a quantum channel and give necessary and sufficient conditions for the channel to have a unique full-rank fixed point. We define the basic distance measures encountered in this thesis (trace norm, Bures, Chernoff, Relative Entropy, χ^2 and Hilbert Metric), and give some useful representations of these measures that are often used implicitly in the remainder of the thesis. We relate the distance measures to each other and discuss their operational interpretations when possible. Having established a proper notion of distance measure, we introduce convergence measure which are meant to estimate the distance to stationarity of the output state of a quantum channel. Finally, we provide a general convergence theorem which characterizes the role of the gap in channel convergence.

Chapter 3:

The goal of this chapter is to introduce some basic notions of quantum Markov chain mixing. We define the mixing time as the number of times a quantum channel has to be applied to an arbitrary initial state before it is close (in trace norm) to its unique stationary state. We generalize one of the basic mixing time bounds from the classical theory of mixing - the so called L^2 bound. In doing so we need to introduce a number of new notions. First, we introduce a quantum generalization of the classical χ^2 divergence measure to the quantum setting. We show that this distance measure does not have a unique generalization, and we provide a family of quantum χ^2 divergence measures which arise naturally in the framework of monotone Riemannian metric on matrix spaces. We prove a number of important properties of the quantum χ^2 -divergence, including monotonicity and joint convexity, and relate its contraction coefficient to the trace norm contraction coefficient.

Having set up an appropriate framework, we discuss a quantum generalization of detailed balance, and we prove a restricted quantum conductance bound (Cheegers inequality) which characterizes the tendency of the quantum channel to exhibit bottlenecks. Our conductance bound holds only for unital channels.

Chapter 4:

In this chapter we introduce the notion of Cutoff Phenomenon, and show that it occurs for the convergence behavior of a tensor product of Markovian channels. Loosely speaking, the cutoff phenomenon describes the situation where the (quantum) Markov chain, for some initial states, stays far away from its stationary distribution for a possibly long time (thus, e.g. preserving classical information), and then, at a specific time that may depend on the system size, suddenly approaches the fixed point (thereby suddenly losing all information from the initial state). Cutoff phenomena depend, apart from the type of Markov chains, on the chosen distance measure. The prototypical example of this behavior is in

card shuffling, where a deck of cards is well shuffled only after a number of shuffles logarithmic in the deck size, whereas before that time it has large dependence on the initial ordering. We state a quantum version of this framework, and apply it to some situations of relevance in Quantum Information Theory.

We show that cutoff type behavior occurs in trace norm for tensor product channels. We sharpen this result by proving that a rigorous cutoff occurs for separable inputs when the channel is primitive, and it always occurs when the stationary state is unique and pure.

Chapter 5:

In this chapter we consider several tasks of dissipative engineering. In the first subsection, we consider the task of dissipatively preparing a maximally entangled state of two atoms in a cavity QED system. We propose a novel scheme for this system such that starting from an arbitrary initial state, a singlet state is prepared as the unique fixed point of a dissipative quantum dynamical process. In our scheme, cavity decay is no longer undesirable, but plays an integral part in the dynamics. As a result, we get a qualitative improvement in the scaling of the fidelity with the cavity parameters. Our analysis indicates that dissipative state preparation is more than just a new conceptual approach, but can allow for significant improvement as compared to preparation protocols based on coherent unitary dynamics.

As a second task of dissipative engineering, we consider the dissipative preparation of graph states. We show that the mixing time scales as $\log n$ in the number of stabilizer elements, and that there exist initial states which indeed saturate this bound.

Finally, as a third task of dissipative engineering, we generalize the clock construction of Kitaev to open systems and show that dissipative computation is equivalent to circuit quantum computation for 5-local Lindblad operators. We rigorously prove rapid convergence of the process.

Chapter 6:

We introduce and apply Hilbert's projective metric in the context of quantum information theory. The metric is induced by convex cones such as the sets of positive, separable or PPT operators. It provides bounds on measures for statistical distinguishability of quantum states and on the decrease of entanglement under LOCC protocols or other cone-preserving operations. The results are formulated in terms of general cones and base norms and lead to contractivity bounds for quantum channels, for instance improving Ruskai's trace-norm contraction inequality. A new duality between distinguishability measures and base norms is provided. For two given pairs of quantum states we show that the contraction of Hilbert's projective metric is necessary and sufficient for the existence of a probabilistic quantum operation that maps one pair onto the other. Inequalities between Hilbert's projective metric and the Chernoff bound, the fidelity and various norms are proven.

Fundamentals

2.1 Basic notions

In this section we define some basic notions of Quantum Information Theory (QIT) which will be used throughout the thesis. This will also help set the notation and conventions straight, which, for the most part, are standard within the QIT community. For a more detailed treatment of the basics of QIT, see [NC00, KSV00], and for good textbooks on Matrix analysis, consult [HJ06, HJ07, Bha97].

We denote an abstract Hilbert space by \mathcal{H} and the set of linear maps on \mathcal{H} by $\mathcal{B}(\mathcal{H})$. Throughout this thesis we will mostly consider linear maps, often denoted T , from the complex d -dimensional matrix algebra \mathcal{M}_d to itself. Note, that \mathcal{M}_d itself becomes a Hilbert space when equipped with the standard Hilbert-Schmidt scalar product $\langle A|B \rangle \equiv \text{tr}[A^\dagger B]$; this Hilbert space is naturally isomorphic to $M_d \simeq \mathbb{C}^{d^2}$. The eigenvalues and singular values of T can be understood in terms of the matrix representation $\hat{T} \in \mathcal{M}_{d^2}$ of T acting on \mathbb{C}^{d^2} . The matrix representation, which will always be written with a hat (ex. \hat{T}), is given in terms of some complete orthonormal basis $\{F_i\}_{i=1\dots d^2}$ of \mathcal{M}_d , where its matrix elements are $\hat{T}_{ij} = \langle F_i|T|F_j \rangle$. Unless otherwise specified, we consider the basis of matrix units. Dual maps T^* can be seen as the hermitian conjugate of T with respect to the Hilbert-Schmidt scalar product. In the above matrix representation of the map, this corresponds exactly to taking the hermitian conjugate \hat{T}^\dagger .

2.1.1 Norms

We will be using a number of different norms on \mathcal{M}_d , most of which can be derived from *Shatten p -norms* (or simply *p -norms*):

$$\|A\|_p := \left(\sum_{i=1}^d [s_i(A)]^p \right)^{1/p}, \quad \text{for any } p \geq 1 \quad (2.1)$$

where $[s_i(A)]_{i=1}^d$ are the singular values of $A \in \mathcal{M}_d$. The Shatten p -norms are unitarily invariant (i.e. $\|UAV\|_p = \|A\|_p$ for all $p \geq 1$ and all unitaries U, V) and have a natural ordering $\|A\|_1 \geq \|A\|_p \geq \|A\|_{p'} \geq \|A\|_\infty$ with $p' \geq p$. Moreover, $p' > p$ leads to a

strict inequality iff A has rank at least two. The case when $p = 1$ is usually referred to as the *trace norm* and the case when $p = \infty$ is called the *operator norm* (or *infinity norm*).

For any $A, B \in \mathcal{M}_d$, unitarily invariant norms in general, and p -norms in particular satisfy a very useful inequality called *Hölder's inequality*: for any $q \geq 1$ and $1/q + 1/r = 1$,

$$\|AB\| \leq \| |A|^q \|^{1/q} \| |B|^r \|^{1/r} \quad (2.2)$$

Applying Hölder's inequality to the trace norm and observing that $|\operatorname{tr}[A^\dagger B]| \leq \operatorname{tr}[|A^\dagger B|]$ leads to an operator Cauchy-Schwarz inequality

$$|\operatorname{tr}[A^\dagger B]| \leq \|A\|_q \|B\|_r \quad (2.3)$$

For $q = r = 2$ this reduces to the Cauchy-Schwarz inequality for the Hilbert-Schmidt inner product. Hölder's inequality also leads to an important set of dimensional dependent inequalities for the p -norms: for $p' \geq p \geq 1$,

$$\|A\|_p \leq d^{1/p - 1/p'} \|A\|_{p'} \quad (2.4)$$

Eqn. (2.4) is usually interpreted as meaning that for finite dimensional spaces all norms are equivalent (up to a dimensional pre-factor). Finally, we mention some important variational characterizations of the p -norms:

Lemma 1. *Let $A, B, P \in \mathcal{M}_d$ with $P \geq 0$ and $1/q + 1/r = 1$. Then*

$$\|A\|_q = \sup\{|\operatorname{tr}[A^\dagger B]| : \|B\|_r = 1\} \quad (2.5)$$

$$\|A\|_1 = \sup\{|\operatorname{tr}[A^\dagger U]| : UU^\dagger = \mathbb{1}\} \quad (2.6)$$

$$\|A\|_\infty = \sup\{|\langle \phi | A | \psi \rangle| : \|\phi\| = \|\psi\| = 1\} \quad (2.7)$$

$$\|P\|_q^q = \frac{1}{q} \sup_{X \geq 0} \left[\operatorname{tr}[PX] - \frac{1}{r} \|X\|_r^r \right] \quad (2.8)$$

The proofs of Lemma 1 and of the previous statements on p -norms can be found in [Bha97].

Superoperator norms

Often times, we will want to work with the norm of a linear map. For this, we introduce an extension of the p -norms to linear maps as

Definition 2. *Let $\Phi : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a linear map, then the $q \rightarrow p$ norm of Φ is*

$$\|\Phi\|_{q \rightarrow p} := \sup_{X \in \mathcal{M}_d, X \neq 0} \frac{\|\Phi(X)\|_p}{\|X\|_q} \quad (2.9)$$

These norms share a number of properties with the Schatten p -norms, with the one outstanding exception that they are not stable under taking tensor products (i.e. $\|\Phi \otimes \Psi\|_{q \rightarrow p} \neq \|\Phi\|_{q \rightarrow p} \|\Psi\|_{q \rightarrow p}$). To remedy this, and in accordance with the requirements of complete positivity, we introduce the stabilized superoperator norms, also called the completely bounded or simply *cb-norms*.

Definition 3. Let $\Phi : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a linear map, then the $q \rightarrow p$ cb-norm of Φ is

$$\|\Phi\|_{cb,q \rightarrow p} := \sup_{X \in \mathcal{M}_d, X \neq 0} \frac{\|\Phi \otimes id_d(X)\|_p}{\|X\|_q}, \quad (2.10)$$

where id_d is the identity map on \mathcal{M}_d .

Below, we list some of the properties of the q, p norms (proofs can be found in [Wat04], and references therein).

- For any linear maps Φ, Ψ , and any $p \geq 1$,

$$\|\Phi \otimes \Psi\|_{cb,1 \rightarrow p} = \|\Phi\|_{cb,1 \rightarrow p} \|\Psi\|_{cb,1 \rightarrow p} \quad (2.11)$$

- For all $p \geq 2$ and $q \leq 2$, and any linear map Φ ,

$$\|\Phi\|_{q \rightarrow p} = \|\Phi\|_{cb,q \rightarrow p} \quad (2.12)$$

- For any completely positive map Φ , the $\|\Phi\|_{q \rightarrow p}$ is reached for a positive state (complete positivity will be defined in the next section).

2.1.2 From classical to quantum probability: quantum Markov chains

A (classical) stochastic system is described by a finite state space Ω consisting of elementary events, or outcomes. A random variable, denoted X, Y, \dots , takes values in Ω^1 . A sequence of random variables X_1, X_2, \dots is called a *Markov chain* if the probability of any given event $x \in \Omega$ at time $n + 1$ depends only on the state of the system at time n , i.e.

$$\mathbb{P}(X_{n+1} = x | X_n = y, X_{n-1} = y', \dots, X_1 = y^{(n-1)}) = \mathbb{P}(X_{n+1} = x | X_n = y) \equiv P_{xy} \quad (2.13)$$

The evolution of the Markov chain can then naturally be described by the *probability transition matrix* P . All elements of $[P_{xy}]_{x,y=1,\dots,|\Omega|}$ are real and non-negative, and $\sum_{x \in \Omega} P_{xy} = 1$ for all $y \in \Omega$. In Matrix analysis, P is said to be *stochastic*. The Markov property is often understood as meaning that the future state of the system does not depend on its history, but only on its state at the present time.

A natural quantum generalization of a probability vector (on L^1) with d outcomes is a density matrix on \mathcal{S}_d . A density matrix ρ is a positive trace-one operator; i.e $\mathcal{S}_d = \{\rho \in \mathcal{M}_d | \rho \geq 0, \text{tr}[\rho] = 1\}$. We will consider a number of restricted sets of density matrices. The set of pure states (rank one density operators) is denoted \mathcal{S}_d^1 , while the set of positive definite states is denoted \mathcal{S}_d^+ . The set of separable states on a specified bipartition $\mathcal{H}_A \otimes \mathcal{H}_B$ will be denoted

$$\mathcal{S}^{\text{sep}} := \{\rho^{AB} \in \mathcal{S}_{d_A d_B} | \rho^{AB} = \sum_k r_k \rho_k^A \otimes \rho_k^B, \rho_k^{A,B} \in \mathcal{S}_{d_{A,B}}, r_k \geq 0, \forall k, \sum_k r_k = 1\} \quad (2.14)$$

¹We formally assume the probability space to be L^1 ; i.e. the states are normalized in one-norm.

A natural quantum extension of the probability transition matrix is a linear completely positive, and trace preserving map (cpt), or *quantum channel*. For every $A, B \in \mathcal{M}_d$ and $\lambda \in \mathbb{C}$, a map $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ is (i) *linear* if $T(A + \lambda B) = T(A) + \lambda T(B)$, (ii) *trace preserving* if $\text{tr}[T(A)] = \text{tr}[A]$, and (iii) *completely positive* if

$$T \otimes \text{id}(AA^\dagger) \geq 0, \quad \forall A \in \mathcal{M}_d \otimes \mathcal{M}_{d'} \quad \text{and} \quad d' \in \mathbb{N} \quad (2.15)$$

The linearity assumption is inherent to quantum mechanics. Trace and positivity preservation guarantee that density matrices are always mapped to density matrices by quantum channels. The assumption of complete positivity might seem peculiar, but it can be seen that positive maps are not sufficient to guarantee that all physical states are again mapped onto physical states. In particular, taking two copies of a system, acting on the first with a quantum channel T and trivially on the second (i.e. $T \otimes \text{id}$) should be a reasonable physical operation. However, there exist positive maps for which this is not the case; where positive inputs on the joint system are mapped onto non-positive output states. The archetypal example of such a map is the transpose on a bipartition [Wer89].

Quantum channels have a number of important representations, here we only mention the *Kraus normal form*, which states that for every quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ there exist a set of at most d^2 non-zero linearly independent *Kraus operators* $\{K_j\}$ such that for any input state $\rho \in \mathcal{S}_d$,

$$T(\rho) = \sum_j K_j \rho K_j^\dagger \quad (2.16)$$

We will also consider time continuous quantum processes, described by a one-parameter semi-group of quantum channels

$$T_t(\rho) = e^{t\mathcal{L}}(\rho) \quad (2.17)$$

The generator of the semi group, the *Liouvillian* \mathcal{L} , has a canonical representation analogous to the Kraus normal form for quantum channels: the *Lindblad normal form*. For every Liouvillian $\mathcal{L} : \mathcal{M}_d \rightarrow \mathcal{M}_d$ there exist a set of at most d^2 non-zero linearly independent *Lindblad operators* $\{L_j\}$ and a Hermitian matrix $H \in \mathcal{M}_d$ (the Hamiltonian) such that for any input state $\rho \in \mathcal{S}_d$,

$$\mathcal{L}(\rho) = i[\rho, H] + \sum_j L_j \rho L_j^\dagger - \frac{1}{2}(L_j^\dagger L_j \rho + \rho L_j^\dagger L_j) \quad (2.18)$$

2.2 Spectral properties of channels

In this section, we go over some of the basic spectral properties of quantum channels. We omit many of the proofs of propositions and refer the interested reader to the set of notes [Wol10], from which many of the statements and notation were borrowed with little modification. For a further source on the spectral properties of quantum channels, consult [WPG10].

Understanding the spectral properties of quantum channels is important for a number of reasons. It is known that the convergence of classical Markov chains is intimately connected with the spectrum of its probability transition matrix, and the same holds true for

quantum channels. In fact, the spectrum of a quantum channel behaves in very much the same way as that of a stochastic matrix. Hence a good understanding of the spectral properties of channels is necessary as a basis for any theory of quantum Markov chain mixing. Furthermore, assumptions about the spectrum of a channel often enter into the statement of convergence theorems. A prime example of this is seen in Chapter 4, where the primitivity of the channel in Proposition 47 and the purity of the fixed point in Proposition 48 allow to prove a cutoff, which we are not able to prove in all generality (Theorem 45). We will first go over the basic spectral properties of quantum channels, and then characterize two special classes of quantum channels (irreducible and primitive).

If a linear map $T : \mathcal{M}_d \rightarrow \mathcal{M}_{d'}$ has equal input and output dimensions ($d = d'$), then we can assign a *spectrum* to it. The *spectrum* of a quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ is given by $\Sigma(T) := \{\lambda | \exists X \in \mathcal{M}_d : T(X) = \lambda X\}$ and its *spectral radius* is $\varrho(T) := \sup\{|\lambda| | \lambda \in \Sigma(T)\}$.

Proposition 4. *Let $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a positive map, then $\varrho(T) \leq \|T(\mathbb{1})\|_\infty$. If T is trace preserving, then $\varrho(T) \leq 1$, implying that $\Sigma(T)$ is contained in the closed unit disk of the complex plain.*

Note that as a consequence of (complete) positivity, a quantum channel T is also hermicity preserving; i.e. $T(X^\dagger) = T(X)^\dagger$. If X is an eigenvector of T with eigenvalue λ , then $\bar{\lambda}X^\dagger = (\lambda X)^\dagger = T(X)^\dagger = \bar{\lambda}T(X^\dagger)$. Hence, X^\dagger is also an eigenvector with eigenvalue $\bar{\lambda}$. In other words, every eigenvalue has a complex conjugate partner.

Observing that the set of Hermitian matrices in \mathcal{M}_d is a real vector space, and that \mathcal{S}_d is a compact, convex set, Brouwer's fixed point theorem [Bro11] implies that every quantum channel has at least one stationary state. In particular, this implies that for any finite dimensional quantum channel T , $\varrho(T) = 1$. Conversely, every eigenvector with eigenvalue one can be written as a complex sum of 4 positive operators, each of which is also a fixed point of T .

The subset of the spectrum consisting of eigenvalues of magnitude 1 ($\Sigma_1 := \{\lambda \in \Sigma(T) | |\lambda| = 1\}$) is called the *peripheral spectrum*.

Proposition 5. *Let $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a quantum channel. If $\lambda \in \Sigma_1$, then all Jordan blocks for λ are one, i.e. its geometric multiplicity equals its algebraic multiplicity.*

Like any linear map, a quantum channel can be decomposed uniquely into Jordan normal form:

$$\hat{T} = \sum_k \mu_k \hat{P}_k + \hat{N}_k ,$$

where each $\mu_k \in \Sigma(T)$ is an eigenvalue of \hat{T} , \hat{P}_k is a projection onto the corresponding Jordan eigenspace, and \hat{N}_k is a nilpotent matrix on this eigenspace. Given that the peripheral spectrum has trivial Jordan blocks ($N_0 = 0$), the asymptotic space of the quantum channel T (also called the *rotating points* of T) is $\aleph_T := \text{span}\{X \in \mathcal{M}_d | \exists \theta \in \mathbb{R} : T(X) = e^{i\theta} X\}$. Its asymptotic evolution is the phase-preserving projection onto \aleph_T ,

$$\hat{T}_\varphi := \sum_{k: |\mu_k|=1} \mu_k \hat{P}_k , \quad (2.19)$$

and the corresponding map $T_\varphi : \mathcal{M}_d \rightarrow \mathcal{M}_d$ is a quantum channel.

If $T_t = e^{t\mathcal{L}}$ comes from a one-parameter semigroup of quantum channels, the spectra and (generalized) eigendecompositions of T_t and of \mathcal{L} are related by exponentiation. In this case, we suppress the time-dependence when writing $T_\varphi := (T_t)_\varphi$.

Note that $(T - T_\varphi)$ has spectral radius equal to $\bar{\mu}$ where $\bar{\mu} := \sup\{|\lambda| \mid \lambda \in \Sigma(T), |\lambda| < 1\}$ is the largest modulus of the eigenvalues of T in the interior of the unit disk. If $T_t = e^{t\mathcal{L}}$ comes from a Markov semigroup, then $\bar{\mu}(t) = e^{-t\bar{\lambda}}$ where $\bar{\lambda} := \inf\{|\operatorname{Re}(\lambda)| \mid \operatorname{Re}(\lambda) < 0, \lambda \in \Sigma(\mathcal{L})\}$, and $\bar{\lambda}$ is referred to as the *gap* of the Liouvillian.

We now give a partial classification of quantum channels, by the nature and properties of their spectra. One situation which is particularly important in the study of convergence rates of Markov channels is when the stationary state of the channel is unique. Classically, this scenario is characterized by the Perron-Frobenius theorem [Per07, Fro12]. We provide here a quantum version of it, and in the process define the notion of an *irreducible* quantum channel.

Proposition 6 (Irreducible channels). *Let $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a quantum channel. The following properties are equivalent:*

1. *Irreducibility: If $P \in \mathcal{M}_d$ is a Hermitian projector satisfying $T(P\mathcal{M}_dP) \subseteq P\mathcal{M}_dP$, then $P \in \{0, \mathbb{1}\}$.*
2. *For every non-zero $A \geq 0$, it holds that $(\operatorname{id} + T)^{d-1}(A) > 0$.*
3. *For every non-zero $A \geq 0$ and every strictly positive $t \in \mathbb{R}$, it holds that $\exp[tT](A) > 0$.*
4. *For every orthogonal pair of non-zero, positive semi-definite $A, B \in \mathcal{M}_d$ there is an integer $t \in \{1, \dots, d-1\}$ such that $\operatorname{tr}[BT^t(A)] > 0$.*

Proof. The proof of Proposition 6 can be found in [EHK78] and references therein \square

If a channel T has only one eigenvalue of modulus one and if its unique stationary state has full rank, then we call T *primitive*. Primitivity, just like irreducibility, allows for a Perron-Frobenius type representation:

Proposition 7 (Primitive channels). *Let $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a quantum channel with Kraus operators $\{K_j\}$. Denote by $\mathcal{K}_m := \operatorname{span}\{\prod_{k=1}^m K_{j_k}\}$ the complex linear span of all degree- m monomials of Kraus operators. Then the following are equivalent:*

1. *T is primitive.*
2. *There exists and $n \in \mathbb{N}$ such that for every non-zero $\psi \in \mathbb{C}^d$ and all $m \geq n$: $\mathcal{K}_m|\psi\rangle = \mathbb{C}^d$.*
3. *There exists a $q \in \mathbb{N}$ such that for all $m > q$: $(T^m \otimes \operatorname{id}_d)(|\Omega\rangle\langle\Omega|) > 0$, where $\Omega \in \mathbb{C}^d \otimes \mathbb{C}^d$ denotes a maximally entangled state.*
4. *There exists a $q \in \mathbb{N}$ such that for all $m > q$: $\mathcal{K}_m = \mathcal{M}_d$.*

Proof. See [SPGWC10] for a proof, and for a detailed discussion of primitive quantum channels. \square

We call a Liouvillian generator \mathcal{L} primitive iff $e^{t\mathcal{L}}$ is a primitive channel for all $t > 0$ or, equivalently, iff $e^{t\mathcal{L}}$ is primitive for one $t > 0$ or, equivalently, iff \mathcal{L} has exactly one eigenvalue 0 and the corresponding stationary state ρ (i.e. $\mathcal{L}(\rho) = 0$) has full rank.

Note: in several of the above theorems, the restriction to quantum channels can be relaxed to Schwarz maps or to positive maps. For simplicity and consistency, we have stated all of the theorems in terms of quantum channels whenever it was possible.

2.2.1 Fixed points of quantum channels

We have already seen some of the properties of the fixed points of quantum channels in the previous section. In particular, we saw that every channel has at least one stationary state. In this section we want to characterize the structure of these stationary states and give conditions for when a channel has a unique fixed point, and when it is of full rank.

Definition 8. Let $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a quantum channel. Then the set of its fixed points will be denoted

$$\mathcal{F}(T) := \{\rho \in \mathcal{S}_d | T(\rho) = \rho\} \quad (2.20)$$

Theorem 9 (Fixed points theorem). Let $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a quantum channel. Then there is a unitary $U \in \mathcal{M}_d$ and a set of positive definite matrices $\rho_k \in \mathcal{S}_{m_k}^+$ such that the fixed point set of T is given by

$$\mathcal{F}(T) = U \left(0 \oplus \bigoplus_{k=1}^K \mathcal{S}_{d_k} \otimes \rho_k \right) U^\dagger \quad (2.21)$$

for an appropriate decomposition of the Hilbert space $\mathbb{C}^d = \mathbb{C}^{d_0} \oplus \bigoplus_{k=1}^K \mathbb{C}^{d_k} \otimes \mathbb{C}^{m_k}$.

Proof. This statement and proof can be traced back to [BKNPV08] and [Lin99]. \square

We now give a condition which guarantees the existence and uniqueness of a full rank fixed point

Theorem 10. Let $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a quantum channel with Kraus decomposition $T(\cdot) = \sum_j K_j \cdot K_j^\dagger$. If for some $n \in \mathbb{N}$, we have

$$\text{span}\left\{\prod_{k=1}^n K_{j_k}\right\} = \mathcal{M}_d, \quad (2.22)$$

i.e. if homogeneous polynomials of the Kraus operators span the entire matrix algebra, then the channel has a unique full rank fixed point.

Corollary 11 (Davies-Frigiero-Spohn criterion). Let $\mathcal{L} : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be the generator of a one-parameter semi-group of quantum channels (a Liouvillian) with standard Lindblad form $\mathcal{L}(\cdot) = i[\cdot, H] + \sum_j L_j \cdot L_j^\dagger - \frac{1}{2}(L_j^\dagger L_j \cdot + \cdot L_j^\dagger L_j)$, then there is a unique full rank fixed point iff

$$\{L_j, L_j^\dagger, H\}' = c\mathbb{1} \quad (2.23)$$

where $\{\cdot\}'$ indicates the commutant, and c is a constant.

Proof. Proofs of this statement were given independently in [Spo77] and [Fri78]. \square

2.3 Distance measures

In order to discuss convergence of a quantum (or classical stochastic) process to its equilibrium distribution, we need to elucidate the notion of distance between quantum states. Distance measures somehow aim at quantifying the overlap between two states. As this notion of "overlap" is quite vague, there is a lot of freedom in the definition. We will see that different measures quantify different aspects of common information content, and we will try to point out the similarities and distinctions between these.

2.3.1 Distance measures review

We recall some of the basic distance measures in quantum information theory, and discuss some of their representations and properties. We introduce two new distance measures - the χ^2 distance and Hilbert's projective metric - which are analyzed in detail in Chapter 3 and 6. For $\rho, \sigma \in \mathcal{S}_d$ consider:

1. The trace distance:

$$d_{\text{tr}}(\rho, \sigma) = \frac{1}{2} \|\rho - \sigma\|_1 \quad (2.24)$$

The trace distance is one of the most widely used distance measures in quantum information theory. Its usefulness comes on the one hand from the fact the it is naturally related to the trace-norm, and on the other hand because it is one of the few distance measures which has a clear operational interpretation. Indeed, as expressed concisely through the Chernoff bound, the trace distance between two quantum states characterizes how distinguishable these two states are if we are allowed to perform arbitrary measurements. This interpretation is seen explicitly through the following variational representation of the trace distance [NC00]:

$$d_{\text{tr}}(\rho, \sigma) = \sup_{\mathbb{1} \geq X \geq 0, X \neq 0} \text{tr}[X(\rho - \sigma)] = \sup_{P=P^2} \text{tr}[P(\rho - \sigma)] \quad (2.25)$$

It turns out that there is another useful variational formula for the trace norm which is a consequence of Eqn. 2.7:

$$d_{\text{tr}}(\rho, \sigma) = \sup\{|\text{tr}[U(\rho - \sigma)]|, UU^\dagger = \mathbb{1}\} \quad (2.26)$$

The trace norm is the tightest distance measure in a sense which will be made precise in the next section. However, it does not distinguish between quantum and classical information, so that its usefulness, in the context of quantum mixing, is mainly in terms of certifying complete information loss.

2. The Bures distance:

$$d_B(\rho, \sigma) = \sqrt{1 - F(\rho, \sigma)} \quad (2.27)$$

where $F(\rho, \sigma) = \text{tr}[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}]$ is the fidelity. The Bures metric also allows for a number of variational characterizations which make it just as convenient to work with as the trace distance. The first of these is the Uhlmann variational representation [Uhl76]:

$$F(\rho, \sigma) = \sup\{\langle \phi, \phi \rangle | \text{tr}_E[|\phi\rangle\langle\phi|] = \rho, \text{tr}_E[|\phi\rangle\langle\phi|] = \sigma\} \quad (2.28)$$

In other words, the fidelity between ρ and σ is given by the largest overlap between two purifications of ρ and σ respectively. This expression is closely related, as the dual semi-definite program [Wat09], to the Alberti variational representation of the fidelity [Alb83]:

$$F(\rho, \sigma) = \inf_{Z \geq 0} \text{tr}[Z\rho] \text{tr}[Z^{-1}\sigma] \quad (2.29)$$

The fidelity can be understood as the natural generalization to mixed states of the inner product between two pure states. The fact that this measure is distinctly quantum mechanical makes it a very appealing measure to use when one wants to consider preservation or loss of quantum information.

3. The Chernoff distance:

$$d_C(\rho, \sigma) = \sqrt{1 - Q(\rho, \sigma)} \quad (2.30)$$

where $Q(\rho, \sigma) = \min_{0 \leq s \leq 1} \text{tr}[\rho^s \sigma^{1-s}]$ is a quantity which appears naturally in the Chernoff bound [ANSV08]. Its logarithm coincides with the asymptotic error rate of quantum hypothesis testing.

4. The relative entropy distance:

$$d_S(\rho, \sigma) = \sqrt{S(\rho, \sigma)} \quad (2.31)$$

where $S(\rho, \sigma) = \text{tr}[\rho(\log(\rho) - \log(\sigma))]$ is the relative entropy. The relative entropy plays a very important role throughout quantum information theory. So much so in fact that no unique operational interpretation can be attributed to it. It is worth mentioning that in the context of asymmetric hypothesis testing [HP91, NO00], the relative entropy replaces the Chernoff function $Q(\rho, \sigma)$ as the exponential of the asymptotic error rate. This distance measure is somehow more closely related to the mutual information content between two state, rather than their distinguishability.

The relative entropy is technically very convenient as it becomes a special limiting case of several families of generalizations: Renyi relative entropies, Tsallis relative entropies, quantum f-divergences, and min-max relative entropies to mention a few. In each of these cases, it is often easier to work with the general family than with the actual relative entropy. Finally we provide a variational characterization of the relative entropy, which is not very well known in the literature [Hia93]:

Lemma 12. *a) If $A \in \mathcal{M}_d$ is Hermitian, and $Y \in \mathcal{M}_d$ $Y > 0$, then*

$$\log \text{tr}[e^{A+\log Y}] = \max\{\text{tr}[\sigma A] - S(X, \sigma) | \sigma \in \mathcal{S}_d^+\} \quad (2.32)$$

b) If $\sigma \in \mathcal{S}_d^+$, and $B \in \mathcal{M}_d$ is Hermitian, then

$$S(\sigma, e^B) = \max\{\text{tr}[\sigma A] - \log \text{tr}[e^{A+B}] | A = A^\dagger\} \quad (2.33)$$

5. The χ^2 -divergence:

$$d_{\chi^2}(\rho, \sigma) = \sqrt{\text{tr}[(\rho - \sigma)\sigma^{-1/2}(\rho - \sigma)\sigma^{-1/2}]} \quad (2.34)$$

The χ^2 -divergence is useful because it can be expressed as an inner product: $d_{\chi^2}^2(\rho, \sigma) = \langle \rho - \sigma, \rho - \sigma \rangle_\sigma$, where $\langle Y, X \rangle_\sigma = \text{tr}[Y^\dagger \sigma^{-1/2} X \sigma^{-1/2}]$. Because of this representation, some mixing properties of a quantum channel can be related to its spectrum in a natural manner. The χ^2 divergence will be analyzed in detail in Chapter 3.

6. Hilbert's projective metric

For $\rho, \sigma \in \mathcal{S}_d$, **Hilbert's projective metric** is given by

$$d_h(\rho, \sigma) = \ln \left[\|\rho^{-1/2} \sigma \rho^{-1/2}\|_\infty \|\sigma^{-1/2} \rho \sigma^{-1/2}\|_\infty \right], \quad (2.35)$$

if $\text{supp}[\rho] = \text{supp}[\sigma]$ and ∞ otherwise.

Hilbert's projective metric is defined on general cones, and can be seen to be invariant under the scaling of its arguments; making it a projective metric. When restricted to a hyperplane of the cone (ex: the trace one hyperplane in the cone of positive operators), then it becomes a proper metric. In Chapter 5, it will be defined precisely, and many of its properties will be worked out.

All of the above distance measures have the following important properties:

- *Non-negativity*: $d(\rho, \sigma) \geq 0$ for all $\rho, \sigma \in \mathcal{S}_d$.
- *Discernability*: $d(\rho, \sigma) = 0$ iff $\rho = \sigma$ for all $\rho, \sigma \in \mathcal{S}_d$.
- *Monotonicity*: For any quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ and any two states $\rho, \sigma \in \mathcal{S}_d$, $d(T(\rho), T(\sigma)) \leq d(\rho, \sigma)$.

The monotonicity property is in particular crucial for their use as convergence measures of a quantum channel. We now provide a partial ordering of these distance measures, by clarifying the relationships between them. We start by relating the trace distance, the Bures distance and the Chernoff distance:

Lemma 13. *For any two states $\rho, \sigma \in \mathcal{S}_d$, the following holds*

$$d_Q^2(\rho, \sigma) \leq d_B^2(\rho, \sigma) \leq d_{\text{tr}}(\rho, \sigma) \leq \sqrt{2} d_B(\rho, \sigma) \leq \sqrt{2} d_Q(\rho, \sigma) \quad (2.36)$$

This relationship can in fact be sharpened when expressed in terms of the fidelity and Q:

$$1 - Q(\rho, \sigma) \leq 1 - F(\rho, \sigma) \leq d_{\text{tr}}(\rho, \sigma) \leq \sqrt{1 - F^2(\rho, \sigma)} \leq \sqrt{1 - Q^2(\rho, \sigma)} \quad (2.37)$$

Proof. The inequalities relating the trace distance and the Bures distance can be found in [NC00], and the inequalities relating the Bures and Chernoff distances were proved in [ANSV08]. \square

It turns out that for states ρ close to a full-rank state $\sigma > 0$, the Bures and trace distances can be bounded even linearly in terms of each other:

Proposition 14 (Trace vs. Bures distance infinitesimally). *Let $\sigma \in \mathcal{S}_d^+$ be a strictly positive density operator with smallest eigenvalue $\lambda_{\min}(\sigma) > 0$. Then there exists $\epsilon = \epsilon(\sigma) > 0$ such that*

$$d_B(\rho, \sigma) \leq \frac{1}{\sqrt{\lambda_{\min}(\sigma)}} d_{\text{tr}}(\rho, \sigma) \quad (2.38)$$

for all density matrices $\rho \in \mathcal{S}_d$ with $d_{\text{tr}}(\rho, \sigma) \leq \epsilon$.

Proof. See Proposition 1 in [KRW11] □

Remark: A general *linear* upper bound of the form $d_B(\rho, \sigma) \leq C d_{\text{tr}}(\rho, \sigma)$ as in Eqn. (2.38) cannot hold for $\sigma \notin \mathcal{S}_d^+$. For instance, in this case there exist density matrices $\rho \in \mathcal{S}_d$ orthogonal to σ ($\rho\sigma = 0$). Then defining $\rho_\delta := \delta\rho + (1 - \delta)\sigma$ (for $\delta \in [0, 1]$) one has $d_{\text{tr}}(\rho_\delta, \sigma) = \delta$ and $F(\rho_\delta, \sigma) = \sqrt{1 - \delta}$, so that

$$d_{\text{tr}}(\rho_\delta, \sigma) \leq 2d_B^2(\rho_\delta, \sigma) \quad \forall \delta \in [0, 1].$$

Thus, there do not exist constants $C, \epsilon > 0$ such that $d_B(\rho, \sigma) \leq C d_{\text{tr}}(\rho, \sigma)$ holds for all $d_{\text{tr}}(\rho, \sigma) < \epsilon$.

We now complete the partial ordering by relating the other distance measures to each other:

Lemma 15. *For any two density operators $\rho, \sigma \in \mathcal{S}_d$, the following hold:*

1. $d_{\text{tr}}(\rho, \sigma) \leq d_{\chi^2}(\rho, \sigma)/2$
2. $d_{\text{tr}}(\rho, \sigma) \leq d_S(\rho, \sigma)/\sqrt{2}$
3. $d_S(\rho, \sigma) \leq d_{\chi^2}(\rho, \sigma)$
4. $d_S^2(\rho, \sigma) \leq d_h(\rho, \sigma)$

Proof. The inequalities relating the trace distance and relative entropy distance to the χ^2 distance were proved in [TKR⁺10] and will be reproduced in the next section, while the inequality relating the trace distance and the relative entropy distance is known as the quantum Pinsker inequality and a proof can be found in [OP04]. The inequality relating the Hilbert distance to the Relative entropy distance was shown in [Dat09]. □

The inequalities above will be used throughout the thesis, and actually reveal a lot of information about what kind of information is accessed in different mixing type bounds.

2.4 Contraction and convergence measures

Having introduced appropriate distance measures between quantum states, we now want to characterize the distance to stationarity of a quantum channel. Such a convergence measure should somehow quantify one or several of the three closely related properties: (i) how far an output state of the channel T may be from its limiting evolution T_φ , (ii) how reversible the action of the channel is, (iii) how much information is lost by the application of the channel. We will now introduce two types of measures of distance to stationarity. The first, which addresses point (i) will be called a *convergence measure* and denoted η_g for some distance measure d_g , whereas the second, which more closely addresses points (ii) and (iii) will be called a *contraction coefficient* and will be denoted with a bar $\bar{\eta}_g$ for some distance measure d_g . We now give formal definitions:

Definition 16 (Convergence and contraction). *Let $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a quantum channel, and let d_g be one of the monotone distance measures introduced in the previous section. Then the g -convergence measure of T is defined as*

$$\eta_g[T] := \sup_{\rho \in \mathcal{S}_d} d_g(T(\rho), T_\varphi(\rho)) \quad (2.39)$$

and the g -contraction coefficient of T is defined as

$$\bar{\eta}_g[T] := \sup_{\rho, \sigma \in \mathcal{S}_d} \frac{d_g(T(\rho), T(\sigma))}{d_g(\rho, \sigma)} \quad (2.40)$$

Both of these measures have their strengths and weaknesses. We point out in particular that $\eta_g[T]$ can be discontinuous in T , whenever the size of the peripheral spectrum is discontinuous, while $\bar{\eta}_g[T]$ is equal to one whenever the channel is not strictly contractive. In particular, whenever $\aleph(T)$ is not one-dimensional (trivial peripheral spectrum), then $\bar{\eta}_g[T] = 1$. We also note, that if the channel T_t is the member of a one-parameter semi-group of quantum channels, and if the peripheral spectrum is trivial, then $\eta_g[T_t]$ and $\bar{\eta}_g[T_t]$ are both finite and strictly smaller than 1 for all $t > 0$. The most widely used distance measure for convergence and contraction is without any doubt the trace distance. The reason for this is that the questions raised often pertain to claims about worst case scenarios, which are well addressed by the trace distance, but also because the trace distance convergence measure and contraction coefficients often allow for easier manipulation. For instance, by convexity of the trace norm, for any quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$,

$$\eta_{\text{tr}}(T) = \sup_{\rho, \sigma \in \mathcal{S}_d} \frac{\|T(\rho - \sigma)\|_1}{\|\rho - \sigma\|_1} = \sup_{\phi, \psi \in \mathcal{S}_d^+, \langle \phi | \psi \rangle = 0} \frac{1}{2} \|T(\psi) - T(\phi)\|_1. \quad (2.41)$$

A proof of this identity can be found in [Rus94]. In particular, if the semigroup is primitive, then we get that:

Proposition 17. *Let $T_t : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be the member of a one-parameter semi-group of primitive quantum channels with $t > 0$, then*

$$\eta_{\text{tr}}[T_t] \leq \bar{\eta}_{\text{tr}}[T_t] \leq 2\eta_{\text{tr}}[T_t] \quad (2.42)$$

Proof. Given that the channel is primitive, it has a unique full rank stationary state; call it $\sigma \in \mathcal{S}_d^+$. This is equivalent to saying that $T_\varphi(\rho) = \sigma$ for any input state $\rho \in \mathcal{S}_d$. We also know that $T_t(\rho)$ will have full support for any input state ρ , whenever $t > 0$. Then,

$$\eta_{\text{tr}}(T) = \frac{1}{2} \sup_{\rho \in \mathcal{S}_d} \|T(\rho) - \sigma\|_1 \quad (2.43)$$

$$\leq \frac{1}{2} \sup_{\rho, \sigma \in \mathcal{S}_d} \|T(\rho) - T(\sigma)\|_1 \quad (2.44)$$

$$= \frac{1}{2} \sup_{\phi, \psi \in \mathcal{S}_1, \langle \phi | \psi \rangle = 0} \|T(\phi - \psi)\|_1 = \bar{\eta}_{\text{tr}}[T] \quad (2.45)$$

The other inequality follows from the triangle inequality and convexity of the trace norm:

$$\bar{\eta}_{\text{tr}}[T] = \frac{1}{2} \sup_{\phi, \psi \in \mathcal{S}_1, \langle \phi | \psi \rangle = 0} \|T(\phi - \psi)\|_1 \quad (2.46)$$

$$\leq \frac{1}{2} \sup_{\phi, \psi \in \mathcal{S}_1, \langle \phi | \psi \rangle = 0} (\|T(\phi) - \sigma\|_1 + \|T(\psi) - \sigma\|_1) \leq 2\eta_{\text{tr}}(T) \quad (2.47)$$

□

We now point out an often very useful relationship between the convergence measure of a channel and the operator norm of the difference between the channel and the projection onto its rotating points.

Proposition 18. *Let $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a quantum channel, and $\|\cdot\|$ the operator norm. Then:*

$$\frac{1}{8\sqrt{d}} \|\hat{T} - \hat{T}_\varphi\| \leq \eta_{\text{tr}}[T] \leq \frac{\sqrt{d}}{2} \|\hat{T} - \hat{T}_\varphi\|. \quad (2.48)$$

Proof. For the lower bound, we use in the first inequality below that every $X \in \mathcal{M}_d$ can be written as $X = P_1 - P_2 + iP_3 - iP_4$ with positive semidefinite P_i satisfying $P_1P_2 = P_3P_4 = 0$, and that then $\|P_j\|_2^2 \leq \sum_i \|P_i\|_2^2 = \|X\|_2^2$. In the following chains we also use that $\|X\|_2 \leq \|X\|_1 \leq \sqrt{d}\|X\|_2$.

$$\begin{aligned} \|\hat{T} - \hat{T}_\varphi\| &= \sup_{\|X\|_2 \leq 1} \|(T - T_\varphi)(X)\|_2 \leq 4 \sup_{P \geq 0, \|P\|_2 \leq 1} \|(T - T_\varphi)(P)\|_2 \\ &\leq 4 \sup_{P \geq 0, \|P\|_1 \leq \sqrt{d}} \|(T - T_\varphi)(P)\|_1 \\ &= 4\sqrt{d} \sup_{P \geq 0, \text{tr}[P] \leq 1} \|(T - T_\varphi)(P)\|_1 = 8\sqrt{d}\eta_{\text{tr}}[T]. \end{aligned} \quad (2.49)$$

$$\begin{aligned} \eta_{\text{tr}}[T] &\leq \frac{1}{2} \sup_{\|X\|_1 \leq 1} \|(T - T_\varphi)(X)\|_1 \\ &\leq \frac{1}{2} \sup_{\|X\|_2 \leq 1} \sqrt{d} \|(T - T_\varphi)(X)\|_2 = \frac{\sqrt{d}}{2} \|\hat{T} - \hat{T}_\varphi\|. \end{aligned} \quad (2.50)$$

□

As the convergence measures are convex in their distance measure, they have the same natural ordering given in Lemmas 13 and 15. The same is not true for the contraction coefficients, as both the numerator and the denominator contribute in determining the supremum. However, the χ^2 and trace norm contraction coefficients are simply related to each other:

Theorem 19. *Given a quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$,*

$$\bar{\eta}_\chi^2(T) \leq \bar{\eta}_{\text{tr}}(T) \leq \bar{\eta}_\chi(T) \quad (2.51)$$

.

Proof. Theorem 19 is proved in Chapter 3, in a slightly more general setting. □

2.5 The Convergence Theorem

We now state an asymptotic convergence theorem for one-parameter semigroups of quantum channels, which is also a crucial fact guiding our further investigations.

Theorem 20 (Contraction theorem). *Let $\mathcal{L} : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a Liouvillian, i.e. the generator of a one-parameter semigroup of quantum channels $T_t \equiv e^{t\mathcal{L}}$ ($t \geq 0$), and let $\bar{\lambda}$ be the gap of \mathcal{L} . Then, there exists $L > 0$ and for any $\nu < \bar{\lambda}$ there exists $R > 0$ such that*

$$Le^{-t\bar{\lambda}} \leq \eta_{\text{tr}}[T_t] \leq Re^{-t\nu} \quad \forall t \geq 0. \quad (2.52)$$

Proof. Let $\hat{\mathcal{L}}$ be the matrix representation of the Liouvillian. The Jordan normal form gives

$$\hat{\mathcal{L}} = \hat{S} \bigoplus_j \hat{J}_j(\lambda_j) \hat{S}^{-1}$$

for some invertible matrix \hat{S} , where λ_j are the eigenvalues of \mathcal{L} , and $\hat{J}_j(\lambda_j)$ are Jordan blocks of the following form (note that eigenvalues λ_j with $\text{Re}(\lambda_j) = 0$ have one-dimensional Jordan blocks, so in particular all 0 eigenvalues):

$$\hat{J}_j(\lambda_j) = \begin{pmatrix} \lambda_j & \lambda_j & 0 & & \\ & \lambda_j & \lambda_j & 0 & \\ & & \lambda_j & \lambda_j & \\ & & & \ddots & \end{pmatrix}.$$

Let $d_j \geq 1$ be the dimension of Jordan block j . Then, in the Jordan basis $\{|k\rangle\}$ defined by this,

$$e^{t\hat{J}_j(\lambda_j)} = e^{t\lambda_j} \sum_{l=1}^{d_j} \sum_{k=1}^l \frac{(t\lambda_j)^{l-k}}{(l-k)!} |k\rangle \langle l|. \quad (2.53)$$

Let $\|\cdot\|$ be the operator norm, and denote by $\kappa(\hat{S}) := \|\hat{S}\| \|\hat{S}^{-1}\|$ the condition number of the similarity transformation into Jordan form. Then:

$$\begin{aligned} \|e^{t\hat{\mathcal{L}}} - \hat{T}_\varphi\| &= \left\| \hat{S} \bigoplus_{j:\text{Re}(\lambda_j) \neq 0} e^{t\hat{J}_j(\lambda_j)} \hat{S}^{-1} \right\| \\ &\leq \kappa(\hat{S}) e^{-t\bar{\lambda}} \max_{j:\text{Re}(\lambda_j) \neq 0} \left\| \sum_{l=1}^{d_j} \sum_{k=1}^l \frac{(t\lambda_j)^{l-k}}{(l-k)!} |k\rangle \langle l| \right\| \\ &\leq \kappa(\hat{S}) e^{-t\bar{\lambda}} \max_{j:\text{Re}(\lambda_j) \neq 0} \left(\sum_{l=1}^{d_j} \sum_{k=1}^l \frac{(t|\lambda_j|)^{l-k}}{(l-k)!} \right) \\ &\leq C e^{-t\bar{\lambda}} \max \{ (t\bar{\lambda})^{J-1}, 1 \}, \end{aligned} \quad (2.54)$$

where C is a t -independent constant and $J := \max_j d_j$ is the dimension of the largest Jordan block. The last step is obtained by factoring $(t\bar{\lambda})^{J-1}$ out of the sum (for times $t \geq 1/\bar{\lambda}$) and bounding the remaining term by a constant. Thus clearly, for any $\nu < \bar{\lambda}$ there exists a constant $K > 0$ such that the last expression is upper bounded by $K e^{-t\bar{\lambda}}$ (for all $t \geq 0$).

The lower bound is obtained similarly (letting $\hat{J}_1(\lambda_1)$ be any Jordan block with $\text{Re}(\lambda_1) = -\bar{\lambda}$):

$$\begin{aligned} \|e^{t\hat{\mathcal{L}}} - \hat{T}_\varphi\| &\geq \kappa(\hat{S})^{-1} \left\| \bigoplus_{j:\text{Re}(\lambda_j) \neq 0} e^{t\hat{J}_j(\lambda_j)} \right\| = \kappa(\hat{S})^{-1} \max_{j:\text{Re}(\lambda_j) \neq 0} \|e^{t\hat{J}_j(\lambda_j)}\| \\ &\geq \kappa(\hat{S})^{-1} \|e^{t\hat{J}_1(\lambda_1)}\| \geq \kappa(\hat{S})^{-1} \max_{1 \leq k \leq l \leq d_1} \left| \left(e^{t\hat{J}_1(\lambda_1)} \right)_{kl} \right| \geq \kappa(\hat{S})^{-1} e^{-t\bar{\lambda}}, \end{aligned}$$

where maximum in the second-to-last expression runs over all matrix elements (k, l) in (2.53), and in the last step we chose $k = l = 1$.

Using these upper and lower bounds on $\|\hat{T}_t - \hat{T}_\varphi\|$, we invoke Lemma 18 (below) to complete the proof of Theorem 20, by lumping all of the time-independent constants together and denoting them by $L > 0$ and R . \square

The proof shows, in particular, that one may not choose $\nu = \bar{\lambda}$ when an eigenvalue λ_k of \mathcal{L} with modulus $|\lambda_k| = \bar{\lambda}$ has a non-trivial Jordan block. Theorem 20 makes a statement only about the asymptotic convergence behavior, i.e. the exponential rate as $t \rightarrow \infty$. If ν is taken close to $\bar{\lambda}$, then R can become very large. But for fixed ν , a universal dimension-dependent upper bound on R of order d^{d^2} can be obtained. If the system at hand is composed of many particles, say n qubits, then this time-independent prefactor can in principle become doubly exponentially large in the number of particles; this would correspond to an exponentially long time to convergence, even for a gap constant in the system size.

Thus, if the phase space of a given process is exponentially large (as is essentially always the case in a many particle system), then even if one can bound the gap, the convergence could still be slow. Hence, in order to prove rapid convergence, it is necessary to bound both ν and R !

Quantum Markov chain mixing

Synopsis:

We introduce quantum versions of the χ^2 -divergence, provide a detailed analysis of their properties, and apply them in the investigation of mixing times of quantum Markov processes. An approach similar to the one presented in [DS91, Fil91, Mih89] for classical Markov chains is taken to bound the trace-distance from the steady state of a quantum processes. A strict spectral bound to the convergence rate can be given for time-discrete as well as for time-continuous quantum Markov processes. Furthermore the contractive behavior of the χ^2 -divergence under the action of a completely positive map is investigated and compared with the contraction of the trace norm. In this context we analyze different versions of quantum detailed balance and, finally, give a geometric conductance bound to the convergence rate for unital quantum Markov processes. Furthermore, we extend the classical concept of detailed balance to the quantum setting and discuss its relevance in general terms.

The chapter is organized as follows; The remainder of section 3.1 is devoted to recalling the framework of classical χ^2 mixing. Then in section 3.2, we introduce the quantum χ^2 -divergence, and prove some basic properties relating it to other divergence and distance measures. In particular we focus on a specific subfamily of interest. In section 3.3, we consider contraction of the χ^2 -divergence under the action of a channel, and relate it to trace-norm contraction. Furthermore, we prove some fundamental quantum mixing time results, whose classical analogues are well known. In section 3.4, we study quantum detailed balance, and in section 3.5, we extend an important classical geometric mixing bound (Cheeger's inequality) to the quantum setting.

Based on:

The χ^2 -divergence and mixing times of quantum Markov processes

K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf and F. Verstraete,
J. Math. Phys. **51**, 122201 (2010).

3.1 Markov chain mixing

The mixing time of a classical Markov chain is the time it takes for the chain to be close to its steady state distribution, starting from an arbitrary initial state. The ability to bound the mixing time is important, for example in the field of computer science, where the bound can be used to give an estimate for the running time of some probabilistic algorithm such as the Monte Carlo algorithm. The mixing time for a classical Markov process P_{ij} , with $\sum_i P_{ij} = 1$ on the space of probability measures \mathcal{S}^{cl} is commonly defined in terms of the one norm, $\|p\|_1 = \sum_i |p_i|$. Let π denote the fixed point of the classical Markov process, i.e. $P\pi = \pi$, then the mixing time is defined as:

$$t_{\text{mix}}(\epsilon) = \min \{n \mid \forall q \in \mathcal{S}, \|P^n q - \pi\|_1 < \epsilon\}. \quad (3.1)$$

A large set of tools has emerged over the years that allows to investigate the convergence rate of classical Markov chains [LPW08]. One of the most prominent approaches to bounding the mixing time of a Markov chain is based on the χ^2 -divergence [Pea00]. This divergence is defined for two probability distributions $p, q \in \mathcal{S}^{\text{cl}}$ as,

$$\chi^2(p, q) = \sum_i \frac{(p_i - q_i)^2}{q_i}. \quad (3.2)$$

The usefulness of the χ^2 -divergence for finding bounds to the mixing time of classical Markov chains arises from the fact that it serves as an upper bound to the one norm difference between two probability distributions,

$$\|p - q\|_1^2 \leq \chi^2(p, q) \quad (3.3)$$

and allows for an easier access to the spectral properties of the Markov chain. The χ^2 -divergence is intimately related to the Kullback-Leibler divergence, or relative entropy, $H(p, q) = \sum_i p_i (\log p_i - \log q_i)$. In fact, it can be obtained directly from the relative entropy as the approximating quadratic form, i.e. as the Hessian, of the latter:

$$\chi^2(p, q) = -\frac{\partial^2}{\partial_\alpha \partial_\beta} H(q + \alpha(p - q), q + \beta(p - q)) \big|_{\alpha=\beta=0}. \quad (3.4)$$

The χ^2 divergence was first introduced by Karl Pearson in the context of statistical inference tests, the most widely used of which is the "Pearson's χ^2 test" [Pea00]. Its computational simplicity and its clear relation to other distance measures have made it one of the most studied divergence measures in the literature.

3.2 The quantum χ^2 -divergence

We want to define a generalization of the classical χ^2 -divergence to the case when we are working on spaces with non-commuting density matrices. We shall require that any generalization to the setting of density matrices satisfies the condition that when the inputs are diagonal, the classical χ^2 -divergence is recovered. The first observation we make, reading straight off from Eqn. (3.2), is that the classical χ^2 -divergence can be seen as an inner product on the probability space weighted with the inversion of the distribution q_i .

Due to the non-commutative nature of density matrices there is no unique generalization of this inversion. Consider for instance a generalization for two density matrices $\rho, \sigma \in \mathcal{S}_d$, where for now we assume σ to be full rank, that is given by

$$\chi_\alpha^2(\rho, \sigma) = \text{tr} \left[(\rho - \sigma) \sigma^{-\alpha} (\rho - \sigma) \sigma^{\alpha-1} \right] = \text{tr} \left[\rho \sigma^{-\alpha} \rho \sigma^{\alpha-1} \right] - 1. \quad (3.5)$$

This gives rise to an entire family of χ^2 -divergences with (as we see below) special properties, for every $\alpha \in [0, 1]$. The natural question of whether there exists a classification of all possible inversions of σ , was investigated in a series of papers by Morozova and Chentsov [MC89] and Petz [Pet96, PS96, PR98], in the context of information geometry. They considered the characterization of monotone Riemannian metrics on matrix spaces. Their general definition is based on the modular operator formalism of Araki [Ara76, Ara87], which we will also consider here. In order to classify the valid inversions, we first need to define the following set of functions, each of which gives rise to a possible inversion:

$$\mathcal{K} = \{k \mid -k \text{ is operator monotone, } k(w^{-1}) = wk(w), \text{ and } k(1) = 1\}. \quad (3.6)$$

Now, we define left and right multiplication operators as $L_Y(X) = YX$ and $R_Y(X) = XY$ respectively. The *modular operator* is defined as

$$\Delta_{\rho, \sigma} = L_\rho R_\sigma^{-1}, \quad (3.7)$$

for all $\rho, \sigma \in \mathcal{S}_d$, $\sigma > 0$. Note, that R_σ and L_ρ commute and inherit hermicity and positivity from ρ, σ . The above should be read as follows: acting on some $A \in M_d$, $\Delta_{\rho, \sigma}(A) = \rho A \sigma^{-1}$. When manipulating the modular operators it is often convenient to write them in matrix form, in which case, they read: $\hat{\Delta}_{\rho, \sigma}|A\rangle = \rho \otimes \bar{\sigma}^{-1}|A\rangle$, where $|A\rangle = A \otimes \mathbb{1}|I\rangle$, and $|I\rangle = \sum_{i=1}^d |ii\rangle$ corresponds to \sqrt{d} times the maximally entangled state. This formalism gives rise to a more general quantum χ^2 -divergence.

Definition 21. For $\rho, \sigma \in \mathcal{S}_d$, and $k \in \mathcal{K}$ we define the quantum χ^2 -divergence

$$\chi_k^2(\rho, \sigma) = \left\langle \rho - \sigma, \Omega_\sigma^k(\rho - \sigma) \right\rangle, \quad (3.8)$$

when $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, and infinity otherwise. The inversion of σ is defined only when $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, and given by

$$\Omega_\sigma^k = R_\sigma^{-1} k(\Delta_{\sigma, \sigma}). \quad (3.9)$$

The functions $k_\alpha(w) = \frac{1}{2} (w^{-\alpha} + w^{\alpha-1})$, with $k_\alpha \in \mathcal{K}$, yield the family of χ_α^2 -divergences given in Eqn. (3.5) which we call the mean α -divergences to distinguish them from the well-known family of WYD α -divergences. Although we focus on the family Eqn. (3.5), most of our results hold for any divergence given by Eqn. (3.8) with $k \in \mathcal{K}$ with the exceptions of Theorem 35.

3.2.1 Monotone Riemannian metrics and generalized relative entropies.

This definition of the χ^2 -divergence stems from the analysis of monotone Riemannian metrics. By Riemannian metric, we mean a positive definite bilinear form $M_\sigma(A, B)$ on the hermitian tangent hyperplane $\mathcal{T}_p = \{A \in \mathcal{M}_d : A = A^\dagger, \text{tr}[A] = 0\}$. The metric is monotone if for all quantum channels $T : \mathcal{M}_d \mapsto \mathcal{M}_d$, states $\sigma \in \mathcal{S}_d^+$ and $A \in \mathcal{T}_p$, $M_{T(\sigma)}(T(A), T(A)) \leq M_\sigma(A, A)$. Petz showed that there is a one-to-one correspondence between the above metrics and a special class of convex operator functions, which correspond to $1/k$ in our notation. He furthermore was able to relate several generalized relative entropies (which he defined much earlier [Pet86] and referred to as quasi-entropies) to monotone Riemannian metrics [PS96, PR98, PG10]. The reverse implication, that every monotone Riemannian metric stems from a generalized relative entropy was first proved by Lesniewski and Ruskai [LR99]. Taking advantage of the well-known integral representations of operator monotone and convex functions [Bha97] one can express the χ^2 -divergences as well as the relative entropies explicitly. We shall briefly repeat the key points of the analysis that are necessary for our understanding of the mixing-time and contraction analysis for cpt-maps.

We need to consider the class of functions \mathcal{G} by which we denote the set of continuous operator convex functions from \mathbb{R}^+ to \mathbb{R} that satisfy $g(1) = 0$. Note that these functions can all be classified in terms of the integral representation:

$$g(w) = a(w-1) + b(w-1)^2 + c \frac{(w-1)^2}{w} + \int_0^\infty \frac{(w-1)^2}{w+s} d\nu(s), \quad (3.10)$$

where $a, b, c > 0$ and the integral of the positive measure $d\nu(s)$ on $(0, \infty)$ is bounded. The generalized relative entropy for states $\rho, \sigma \in \mathcal{S}_d^+$ was first defined in [Pet86].

Definition 22. Let $g \in \mathcal{G}$. The generalized quantum relative entropy is given by

$$H_g(\rho, \sigma) = \text{tr} [\rho^{1/2} g(\Delta_{\sigma, \rho}) (\rho^{1/2})] \quad (3.11)$$

when $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, and infinity otherwise, and where $\Delta_{\rho, \sigma}$ is again the modular operator.

We now recall without proof a theorem [PS96, PR98, LR99] relating the relative entropy and the monotone Riemannian metric, mirroring the classical result Eqn. (3.4):

Theorem 23. For every $k \in \mathcal{K}$, there is a $g \in \mathcal{G}$ such that for a given $\sigma \in \mathcal{S}_d$, and A, B hermitian traceless, we get:

$$\begin{aligned} M_\sigma^k(A, B) &= - \frac{\partial^2}{\partial \alpha \partial \beta} H_g(\sigma + \alpha A, \sigma + \beta B) \Big|_{\alpha=\beta=0} \\ &= \left\langle A, \Omega_\sigma^k(B) \right\rangle. \end{aligned} \quad (3.12)$$

and, k is related to g by

$$k(w) = \frac{g(w) + wg(w^{-1})}{(w-1)^2} \quad (3.13)$$

From this theorem follows a convenient integral representation of the inversion Ω_σ^k , which is equivalent to Eqn. (3.9) [LR99].

$$\Omega_\sigma^k = \int_0^\infty \left(\frac{1}{sR_\sigma + L_\sigma} + \frac{1}{R_\sigma + sL_\sigma} \right) N_g(s) ds, \quad (3.14)$$

where N_g denotes the singular measure $N_g(s)ds = (b_g + c_g)\delta(s)ds + d\nu_g(s)$. Note, that the relationship between k and g is not one-to-one. Indeed, by setting $\hat{g}(w) = wg(w^{-1})$, we get back the above relation. However, there is a one-to-one correspondence between each k and a symmetric $g_s(w) = g(w) + wg(w^{-1})$, and hence between each metric and a symmetric relative entropy.

Note that the α -subfamily of Eqn. (3.5) has the associated symmetric relative entropy: $g^{sym}(x) = \frac{(1-w)^2}{2} (w^{\alpha-1} + w^{-\alpha})$, so that

$$H_\alpha^{sym}(\rho, \sigma) = \frac{1}{2} (H_\alpha(\rho, \sigma) + H_\alpha(\sigma, \rho)) \quad (3.15)$$

where,

$$H_\alpha(\rho, \sigma) = \text{tr} [\rho^{2-\alpha} \sigma^{\alpha-1} + \rho^{1+\alpha} \sigma^{-\alpha} - 2\rho^\alpha \sigma^{1-\alpha}].$$

The integral representation Eqn. (3.14) of the inversion Ω_σ^k allows for a partial ordering of different monotone Riemannian metrics that follows from the set of inequalities:

$$\frac{2}{x+1} \leq \frac{1+s}{2} \left(\frac{1}{s+x} + \frac{1}{sx+1} \right) \leq \frac{x+1}{2x}. \quad (3.16)$$

for $s \in [0, 1]$, and $x \in \mathbb{R}^+$. We therefore see that there exists a partial ordering for the inversions, with a lowest and highest element in the hierarchy. The lowest element gives rise to the so called Bures metric. Thus,

$$\Omega_\sigma^{Bures} = 2(R_\sigma + L_\sigma)^{-1} \leq \Omega_\sigma^k \leq (L_\sigma^{-1} + R_\sigma^{-1})/2 = \Omega_\sigma^{\alpha=0} \quad (3.17)$$

The χ^2 -divergence is recovered from the metric upon setting $\chi_k^2(\rho, \sigma) \equiv M_\sigma^k(\rho - \sigma, \rho - \sigma)$. We are therefore left with a partial order for all possible χ^2 -divergences with a smallest and largest element according to,

$$\chi_{Bures}^2(\rho, \sigma) \leq \chi_k^2(\rho, \sigma) \leq \chi_{\alpha=0}^2(\rho, \sigma). \quad (3.18)$$

The defining attribute of the above set of metrics is their monotonicity under the action of quantum channels. This was first shown by Petz [PR98], and later a proof based on the integral representation of Ω_σ^k , see Eqn. (3.14), and on Schwarz-type inequalities, was provided by Ruskai and Lesniewski in [LR99]. Due to its importance for the mixing time analysis we shall repeat it here.

Theorem 24. *For all $\sigma \in \mathcal{S}_d$, M_σ^k is monotone under the action of a quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ for all $k \in \mathcal{K}$ and $A \in \mathcal{M}_d$, i.e.*

$$M_\sigma^k(A, A) \geq M_{T(\sigma)}^k(T(A), T(A)) \quad (3.19)$$

Proof. The monotonicity follows immediately from the integral representation of the inversion Ω_σ^k in Eqn. (3.14), and an argument proved below; Lemma 25. \square

The proof of the contractivity of a general Riemannian metric is based on the following lemma first proved in [LR99].

Lemma 25. *For a channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$, we have that,*

$$\begin{aligned} \operatorname{tr} \left[A^\dagger \frac{1}{R_\sigma + sL_\rho} A \right] &= \operatorname{tr} \left[T \left(A^\dagger \frac{1}{R_\sigma + sL_\rho} A \right) \right] \geq \\ &\operatorname{tr} \left[T(A)^\dagger \frac{1}{R_{T(\sigma)} + sL_{T(\rho)}} T(A) \right]. \end{aligned} \quad (3.20)$$

Proof. Let $\sigma > 0$, then $\operatorname{tr}[A^\dagger \sigma A] \geq 0$, and $\operatorname{tr}[A^\dagger A \sigma] \geq 0$ so that L_σ as well as R_σ are both positive semi definite super operators on the matrix space. Therefore we infer, that for a positive $\rho > 0$ the operator $R_\sigma + sL_\rho$ is also positive. We define a matrix $X = [R_\sigma + sL_\rho]^{-1/2}(A) + [R_\sigma + sL_\rho]^{1/2}T^*(A)$ and furthermore $B = [R_{T(\sigma)} + sL_{T(\rho)}]^{-1}T(A)$. Since $\operatorname{tr}[X^\dagger X] \geq 0$, we have that

$$\operatorname{tr} \left[A^\dagger \frac{1}{R_\sigma + sL_\rho} A \right] - \operatorname{tr} [T^*(B^\dagger)A] - \operatorname{tr} [A^\dagger T^*(B)] + \operatorname{tr} [T^*(B^\dagger)[R_\sigma + sL_\rho]T^*(B)] \geq 0. \quad (3.21)$$

Furthermore note, that

$$-\operatorname{tr} [A^\dagger T^*(B)] - \operatorname{tr} [T^*(B^\dagger)A] = -2 \operatorname{tr} \left[T(A^\dagger) \frac{1}{R_{T(\sigma)} + sL_{T(\rho)}} T(A) \right]. \quad (3.22)$$

It therefore suffices to show that we are able to bound the last term in Eqn. (3.21) by the right side of the inequality in Eqn. (3.20). Note, that

$$\begin{aligned} \operatorname{tr} [T^*(B^\dagger)[R_\sigma + sL_\rho]T^*(B)] &= \operatorname{tr} [T^*(B^\dagger)T^*(B)\sigma + sT^*(B^\dagger)\rho T^*(B)] \\ &\leq \operatorname{tr} [T^*(B^\dagger B)\sigma + sT^*(BB^\dagger)\rho], \end{aligned} \quad (3.23)$$

since $\rho, \sigma > 0$ and due to the operator inequality $T^*(B^\dagger)T^*(B) \leq T^*(B^\dagger B)$. This inequality holds for any B since T is a channel and by that trace preserving, hence $T^*(\mathbb{1}) = \mathbb{1}$. With $\operatorname{tr} [T^*(B^\dagger B)\sigma] = \operatorname{tr} [B^\dagger B T(\sigma)]$ we can write

$$\begin{aligned} \operatorname{tr} [T^*(B^\dagger)[R_\sigma + sL_\rho]T^*(B)] &\leq \operatorname{tr} [B^\dagger B T(\sigma) + sB^\dagger B T(\rho)] \\ &= \operatorname{tr} [B^\dagger [R_{T(\sigma)} + sL_{T(\rho)}]B] = \operatorname{tr} [B^\dagger T(A)] = \operatorname{tr} \left[T(A^\dagger) \frac{1}{R_{T(\sigma)} + sL_{T(\rho)}} T(A) \right]. \end{aligned} \quad (3.24)$$

□

3.2.2 Properties of the quantum χ^2 -divergence

The fact that the quantum χ_k^2 -divergence can be used to bound the mixing time lies in the following Lemma, that upper bounds the trace distance which is the relevant distance measure in the mixing time definition.

Lemma 26. *For every pair of density operators $\rho, \sigma \in S_d$, we have that*

$$\|\rho - \sigma\|_1^2 \leq \chi_k^2(\rho, \sigma) \quad (3.25)$$

Proof. If the support of ρ is not contained in the support of σ , then the right hand side is ∞ . We can therefore assume w.l.o.g. that $\sigma > 0$ by restricting the analysis to the support space of σ . The trace norm $\|A\|_1$ of some matrix $A \in \mathcal{M}_d$ can be expressed as $\|A\|_1 = \max_{U \in U(d)} \text{tr}[UA]$, where the maximum is taken over all unitaries acting on the d -dimensional Hilbert space. Thus, for any inversion Ω_σ^k :

$$\begin{aligned} \|A\|_1^2 &= \max_{U \in U(d)} \text{tr}[UA]^2 = \max_{U \in U(d)} \text{tr} \left[U[\Omega_\sigma^k]^{-1/2} \circ [\Omega_\sigma^k]^{1/2}(A) \right]^2 \\ &= \max_{U \in U(d)} \text{tr} \left[[\Omega_\sigma^k]^{-1/2}(U) [\Omega_\sigma^k]^{1/2}(A) \right]^2 \\ &\leq \text{tr} \left[A^\dagger \Omega_\sigma^k(A) \right] \max_{U \in U(d)} \text{tr} \left[U^\dagger [\Omega_\sigma^k]^{-1}(U) \right] \end{aligned} \quad (3.26)$$

Let us consider the Bures inversion given by $\Omega_\sigma^{\text{Bures}} = 2[L_\sigma + R_\sigma]^{-1}$. Clearly, its inverse is $[\Omega_\sigma^{\text{Bures}}]^{-1} = \frac{1}{2}[L_\sigma + R_\sigma]$. Therefore, for any unitary U ,

$$\text{tr} \left[U^\dagger [\Omega_\sigma^{\text{Bures}}]^{-1}(U) \right] = \frac{1}{2} \left(\text{tr}[U^\dagger \sigma U] + \text{tr}[U^\dagger U \sigma] \right) = 1. \quad (3.27)$$

Setting $A = \rho - \sigma$ and observing that $\chi_{\text{Bures}}^2 \leq \chi_k^2$ for all $k \in \mathcal{K}$ completes the proof. \square

We have already stated that the family of χ_α^2 -divergences defined in Eqn. (3.5) can be cast into the general framework of monotone Riemannian metrics. Because of its computational simplicity, and its special symmetry when $\alpha = 1/2$, we consider its properties more specifically. It is possible for instance to show monotonicity of this subfamily using arguments from joint convexity. As the proof is interesting in its own right, we give it here:

Proposition 27. χ_α^2 is jointly convex in its arguments for $\alpha \in [0, 1]$. Moreover, it is monotone w.r.t. completely positive trace-preserving maps, i.e.,

$$\chi_\alpha^2(\rho, \sigma) \geq \chi_\alpha^2(T(\rho), T(\sigma)), \quad (3.28)$$

for every quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$.

Proof. A direct application of Corollary 2.1 in [Lie73] guarantees that $\chi_\alpha^2(\rho, \sigma)$ is jointly convex in its arguments for any $\alpha \in [0, 1]$. This in turn implies monotonicity w.r.t. cp-maps by a standard argument: let us represent T as $T(\rho) = \text{tr}_E [U(\rho \otimes \psi)U^\dagger]$ where ψ is a pure state (i.e. rank-one projection), U a unitary and tr_E the partial trace over an ‘environmental’ system of dimension m . If we take a unitary operator basis $\{V_i\}_{i=1, \dots, m^2}$ in \mathcal{M}_m (orthonormal w.r.t. the Hilbert-Schmidt inner product), we can write

$$T(\rho) \otimes \mathbb{1}_m / m = \frac{1}{m^2} \sum_{i=1}^{m^2} (\mathbb{1} \otimes V_i) U(\rho \otimes \psi) U^\dagger (\mathbb{1} \otimes V_i^\dagger). \quad (3.29)$$

However, since $\chi_\alpha^2(T(\rho), T(\sigma)) = \chi_\alpha^2(T(\rho) \otimes \tau, T(\sigma) \otimes \tau)$ in particular for $\tau = \mathbb{1}_m / m$, we can now apply joint convexity. With the help of the fact that for any unitary W it holds that $(W \cdot W^\dagger)^\alpha = W(\cdot)^\alpha W^\dagger$ we obtain the claimed result. \square

Furthermore, we note that this subfamily also has a natural ordering.

Proposition 28. *For every $\rho, \sigma \in \mathcal{S}_d$, χ_α^2 is convex in α , and reaches a minimum for $\alpha = 1/2$.*

Proof. First note that $\chi_{\alpha=0}^2(\rho, \sigma) = \chi_{\alpha=1}^2(\rho, \sigma)$. That the minimum is reached for $\alpha = 1/2$ follows directly from the Cauchy-Schwarz inequality. Applied to our problem we get

$$\begin{aligned} \text{tr} \left[\rho \sigma^{-1/2} \rho \sigma^{-1/2} \right]^2 &= \text{tr} \left[\rho \sigma^{(\alpha-1)/2} \sigma^{-\alpha/2} \rho \sigma^{(\alpha-1)/2} \sigma^{-\alpha/2} \right]^2 \\ &\leq \text{tr} \left[\rho \sigma^{-\alpha} \rho \sigma^{\alpha-1} \right]^2 \end{aligned} \quad (3.30)$$

To see convexity, consider the second partial derivative of χ_α^2 with respect to α :

$$\begin{aligned} \frac{\partial^2}{\partial \alpha^2} \chi_\alpha^2(\rho, \sigma) &= \text{tr} \sigma^{\alpha-1} \rho \sigma^{-\alpha} (\rho \log^2 \sigma + \log^2 \sigma \rho - 2 \log \sigma \rho \log \sigma) \\ &= \sum_{kl} \mu_k^{\alpha-1} \mu_l^{-\alpha} (\log \mu_k - \log \mu_l)^2 |\langle k | \rho | l \rangle|^2 \geq 0 \end{aligned} \quad (3.31)$$

where we used $\sigma = \sum_k \mu_k |k\rangle \langle k|$. □

Finally, we point out a bound on the relative entropy in terms of the α -subfamily of χ^2 -divergences:

Theorem 29 (Relative Entropy bound). *For every pair of density operators ρ and σ and every $\alpha \in (0, 1]$ we have that*

$$\chi_\alpha^2(\rho, \sigma) \geq S(\rho, \sigma), \quad (3.32)$$

where $S(\rho, \sigma) = \text{tr} \rho (\log \rho - \log \sigma)$ is the usual relative entropy.

Proof. It was shown in [RS90] that for $\gamma \in (0, 1]$, the following holds:

$$S(\rho, \sigma) \leq \frac{1}{\gamma} (\text{tr} \rho^{1+\gamma} \sigma^{-\gamma} - 1) \quad (3.33)$$

Then consider,

$$\begin{aligned} \chi_\alpha^2(\rho, \sigma) - S(\rho, \sigma) &\geq \text{tr} \rho \sigma^{-1/2} \rho \sigma^{-1/2} - 2 \text{tr} \rho^{3/2} \sigma^{-1/2} + 1 \\ &= \text{tr} (\rho^{1/2} \sigma^{-1/2} \rho^{1/2} - \rho^{1/2})^2 \geq 0 \end{aligned} \quad (3.34)$$

where the first inequality comes from taking $\gamma = 1/2$ in Eqn. (3.2.2), and $\alpha = 1/2$ for χ_α^2 , and the last line is obtained from rearranging terms. □

3.3 Mixing time bounds and contraction of the χ^2 -divergence.

3.3.1 Mixing time Bounds

The χ^2 -divergence is an essential tool in the study of Markov chain mixing times, because on the one hand it bounds the trace distance, and on the other it allows easy access to the spectral properties of the map. The subsequent analysis can be seen as a generalization of the work presented in [DS91, Fil91] to the non-commutative setting.

Theorem 30 (Mixing time bound). *Let $T : \mathcal{M}_d \mapsto \mathcal{M}_d$ be a primitive quantum channel with fixed point $\sigma \in \mathcal{S}_d^+$, for any $\rho \in \mathcal{S}_d$ and any $k \in \mathcal{K}$, we can bound*

$$\|T^n(\rho) - \sigma\|_{\text{tr}} \leq (s_1^k)^n \sqrt{\chi_k^2(\rho, \sigma)}. \quad (3.35)$$

Here s_1^k denotes the second largest singular value (the largest being 1) of the map

$$Q_k = [\Omega_\sigma^k]^{1/2} \circ T \circ [\Omega_\sigma^k]^{-1/2} \quad (3.36)$$

Before we prove Theorem 30, we would like to point out an important fact that regards the singular values of Q_k . The monotonicity of the χ^2 -divergence ensures, that the singular values s_i^k of Q_k are always contained in $[0, 1]$ irrespectively of the choice of $k \in \mathcal{K}$. Let us therefore prove the following:

Lemma 31 (spectral interval). *The spectrum of the map $S_k \equiv Q_k^* \circ Q_k = [\Omega_\sigma^k]^{-1/2} \circ T^* \circ \Omega_\sigma^k \circ T \circ [\Omega_\sigma^k]^{-1/2}$ is contained in $[0, 1]$.*

Proof. Let us first note, that the map S_k is hermitian and positive by construction. Furthermore, the monotonicity of the χ^2 -divergence, as stated in Theorem 24 ensures that the Rayleigh-Ritz quotient is bounded by 1. This holds, since $\forall B$

$$\begin{aligned} \langle B, S_k(B) \rangle &= \langle A, T^* \circ \Omega_\sigma^k \circ T(A) \rangle = M_{T(\sigma)}^k(T(A), T(A)) \leq \\ M_{\sigma}^k(A, A) &= \langle A, \Omega_\sigma^k(A) \rangle = \langle B, B \rangle, \end{aligned} \quad (3.37)$$

where we defined the intermediate state $A = [\Omega_\sigma^k]^{-1/2}(B)$. Note that we made use of the fact that $\sigma = T(\sigma)$ is the fixed point of the map. Therefore

$$\lambda_{\max} = \max_{B \in \mathcal{M}_d} \frac{\langle B, S_k(B) \rangle}{\langle B, B \rangle} \leq 1 \quad (3.38)$$

and the maximum is attained for $\lambda_{\max} = 1$ and $B_{\max} = [\Omega_\sigma^k]^{1/2}(\sigma)$. \square

With the bound on the spectrum at hand, it is now straight forward to prove Theorem 30

Proof. Define $e(n) \in \mathcal{M}_d$, as $e(n) = T^n(\rho - \sigma)$. By Lemma 26, we get $\|e(n)\|_1^2 \leq \chi_k^2(T^n(\rho), T^n(\sigma)) \equiv \chi_k^2(n)$. In the matrix representation, $|e(n)\rangle = e(n) \otimes \mathbb{1}|I\rangle$, we can rewrite $\chi_k^2(n) = \langle e(n) | \hat{\Omega}_\sigma^k | e(n) \rangle$. Note that also, $|e(n+1)\rangle = \hat{T}|e(n)\rangle$ and so,

$$\chi_k^2(n) - \chi_k^2(n+1) = \langle e(n) | \hat{\Omega}_\sigma^k | e(n) \rangle - \langle e(n) | \hat{T}^\dagger \hat{\Omega}_\sigma^k \hat{T} | e(n) \rangle \quad (3.39)$$

$$= \langle e(n) | [\hat{\Omega}_\sigma^k]^{1/2} \left(\mathbb{1} - \hat{Q}_k^\dagger \hat{Q}_k \right) [\hat{\Omega}_\sigma^k]^{1/2} | e(n) \rangle. \quad (3.40)$$

Due to Lemma 31 we know that the spectrum of $\hat{S}_k = \hat{Q}_k^\dagger \hat{Q}_k$, which is equal to the square of the singular values of \hat{Q}_k , is contained in the interval $[0, 1]$. Hence,

$$\langle e(n) | [\hat{\Omega}_\sigma^k]^{1/2} (\mathbb{1} - \hat{S}_k) [\hat{\Omega}_\sigma^k]^{1/2} | e(n) \rangle \quad (3.41)$$

$$\geq (1 - s_1^2) \langle e(n) | [\hat{\Omega}_\sigma^k]^{1/2} \sum_{\alpha \neq 0} P_\alpha [\hat{\Omega}_\sigma^k]^{1/2} | e(n) \rangle. \quad (3.42)$$

The sum is taken over spectral projectors P_α^k of $\hat{S}_k = \sum_\alpha (s_\alpha^k)^2 P_\alpha$, apart from P_0^k which projects onto $[\hat{\Omega}_\sigma^k]^{-1/2} |\sigma\rangle$. In particular, $P_0^k = [\hat{\Omega}_\sigma^k]^{-1/2} |\sigma\rangle \langle \sigma| [\hat{\Omega}_\sigma^k]^{-1/2}$, so that

$$\langle e(n) | [\hat{\Omega}_\sigma^k]^{1/2} P_0^k [\hat{\Omega}_\sigma^k]^{1/2} | e(n) \rangle = \langle e(n) | \sigma \rangle \text{tr} [T^n(\rho - \sigma)] = 0, \quad (3.43)$$

by trace preservation of T . We can write,

$$\chi_k^2(n) - \chi_k^2(n+1) \geq (1 - (s_1^k)^2) \chi_k^2(n). \quad (3.44)$$

Rearranging terms completes the theorem. \square

Remark: The fact, that the singular values of Q_k are always smaller or equal to one justifies the use of the generalized χ^2 -divergence as the appropriate distance measure to bound the convergence of an arbitrary channel. It is tempting to use the Hilbert-Schmidt inner product to give an upper bound to the trace norm. This can always be done at the cost of a dimension dependent prefactor, since on finite dimensional spaces all norms are equivalent. However, when doing so a problem arises if one tries to bound the convergence in terms of the spectral properties of the map $S_{HS} = T^* \circ T$. It is in general not ensured that the spectrum will be bounded by one. In fact, for every non-unital channel T , S_{HS} will have an eigenvalue larger than one [PGWPR06]. The similarity transformation of the channel T with $[\Omega_\sigma^k]^{1/2}$ alters the singular values, but of course leaves the spectrum invariant. Furthermore, it is a well known fact [HJ07] that the singular values of a square matrix log-majorize the absolute value of the eigenvalues. As the spectrum of Q_k is bounded by one (and equal that of \hat{T} by similarity), we conclude that its second largest eigenvalue is always smaller or equal to its second largest singular value. One can also give a general bound in terms of the second largest eigenvalue of T (cf. Theorem 20 and subsequent comments), but one is then confronted with a potentially severe dimensional prefactor.

For some instances of the inversion Ω_σ^k it becomes immediately evident that the symmetrization S_k has the desired spectral properties without making use of the monotonicity of the χ_k^2 -divergence. It can occur, that S_k is again similar to a quantum channel that is of the form $T_s^k = [\Omega_\sigma^k]^{-1/2} \circ S_k \circ [\Omega_\sigma^k]^{1/2}$. A possible example of such an inversions is $\Omega_\sigma^{\alpha=1/2} = L_\sigma^{-1/2} R_\sigma^{-1/2}$. This is however not the generic case, most inversions will lead to maps that are not completely positive any longer. It would be very desirable to find other such examples, as they mirror the classical situation where the symmetrized maps are always probability transition matrices, and because these specific inversions allow for clean contraction bounds as will be seen in section 3.3.2.

It is clear from the discussion above that the singular values of Q_k play a crucial role in the mixing time analysis presented here. This seems to contradict the general understanding that the convergence is determined by the spectral properties of the channel T in the asymptotic limit. This can however be understood as follows: the matrix \hat{Q}_k is similar to \hat{T} , i.e. $\hat{Q}_k = [\hat{\Omega}_\sigma^k]^{1/2} \cdot \hat{T} \cdot [\hat{\Omega}_\sigma^k]^{-1/2}$, so the spectra of Q_k and T coincide. The following lemma establishes a relation between the singular values and the eigenvalues in the asymptotic limit. For a proof, see e.g. [HJ06] pg.180.

Lemma 32 (Singular values). *Let $\hat{Q}_k \in \mathcal{M}_{d^2}$ be given, and let $s_0(\hat{Q}_k) \geq \dots \geq s_{d^2-1}(\hat{Q}_k)$ and $\{\lambda_i(\hat{Q}_k)\}_{i=0 \dots d^2-1}$ denote its singular values and eigenvalues, respectively with $|\lambda_0(\hat{Q}_k)| \geq \dots \geq |\lambda_{d^2-1}(\hat{Q}_k)|$. Then*

$$\lim_{n \rightarrow \infty} [s_i(\hat{Q}_k^n)]^{1/n} = |\lambda_i(\hat{Q}_k)| \quad \forall i = 0 \dots d^2 - 1 \quad (3.45)$$

In the limit of $n \rightarrow \infty$ applications of the quantum channel, we can start blocking the channel in m subsequent applications $T^{(m)} \equiv T^m$ and bound the convergence rate as a function of the singular values of the corresponding $\hat{Q}_k^{(m)}$, which indeed converge to the eigenvalues of the original cp-map. Convergence following the eigenvalue is therefore only guaranteed in the limit of $n \rightarrow \infty$, and this would indeed be the case, when e.g. the eigenstructure of the original cp-maps contains a Jordan block associated to the second largest eigenvalue. Note, that convergence in the above lemma goes typically as $1/n$, which is very slow. Hence for finite n , convergence is governed by the singular values of \hat{Q}_k as opposed to the eigenvalues. The bound derived in Eqn. (30) is an absolute bound for finite n and clearly leads to a strictly monotonic decay. Note that in the case that the second largest singular value is also equal to 1, this can then always be cured by blocking the cp-maps together. Finally, it is worth mentioning that the convergence can in fact be much more rapid if one starts in a state "closer" to the fixed point. In particular, if the initial state is such that $\rho - \sigma \propto Y_k$, $k \geq 2$, where Y_k is the eigenvector corresponding to λ_k , then the convergence will be governed by the magnitude of λ_k . Furthermore, if instead of a single fixed point, we have a fixed subspace, or a collection of fixed subspaces (with or without rotating points), then the convergence to this fixed subspace will be governed by the largest eigenvalue whose magnitude is strictly smaller than one.

Thus far we have only considered the time-discrete case, it is however straightforward to give a similar bound for time-continuous Markov processes, that are described by a one parameter semi-group. The following lemma bounds the trace-distance as a function of $t \in \mathbb{R}_0^+$: The proof of the following lemma is very similar to the proof of the time discrete case, we will therefore omit it here.

Lemma 33 (Time-continuous bound). *Let \mathcal{L} denote the generator of a one-parameter semigroup of primitive quantum channels, described by the master equation $\partial_t \rho = \mathcal{L}(\rho)$, with solution $\rho(t) \in \mathcal{S}_d \forall t \in [0, +\infty)$. Furthermore let $\sigma \in \mathcal{S}_d^+$ denote the fixed-point $\mathcal{L}(\sigma) = 0$, then*

$$\|\rho(t) - \sigma\|_{\text{tr}}^2 \leq e^{l_1^k t} \chi_k^2(\rho(0), \sigma). \quad (3.46)$$

Here, $l_1^k \leq 0$ refers to the second largest eigenvalue of

$$\Lambda_k = [\Omega_\sigma^k]^{1/2} \circ \mathcal{L}^* \circ [\Omega_\sigma^k]^{-1/2} + [\Omega_\sigma^k]^{-1/2} \circ \mathcal{L} \circ [\Omega_\sigma^k]^{1/2}. \quad (3.47)$$

The symmetrization for the generator of the time continuous Markov process is additive as would be expected. Furthermore, we note that the monotonicity of the χ^2 -divergence ensures that the spectrum of Λ_k is never positive, based on a similar reasoning as given in Lemma (31).

3.3.2 Contraction Coefficients

In the following we study the contraction of the χ^2 -divergences under quantum channels, and its relation to the trace norm contraction. We consider general contraction rather than contraction to the fixed point because analytic results are more readily available, and because these bounds are in a sense the *most* stringent one can require. We focus primarily on the mean α -subfamily of χ^2 -divergences.

Let us recall the definitions of the following contraction coefficients (χ^2 - and trace norm-contraction cf. Eqn. (2.40)):

$$\bar{\eta}_\chi^\alpha(T) = \sup_{\rho, \sigma \in \mathcal{S}_d} \frac{\chi_\alpha^2(T(\rho), T(\sigma))}{\chi_\alpha^2(\rho, \sigma)} \quad (3.48)$$

and

$$\bar{\eta}_{\text{tr}}(T) = \sup_{\rho, \sigma \in \mathcal{S}_d} \frac{\|T(\rho - \sigma)\|_1}{\|\rho - \sigma\|_1} = \sup_{\phi, \psi \in \mathcal{S}_d^1, \langle \phi | \psi \rangle = 0} \frac{1}{2} \|T(\psi) - T(\phi)\|_1, \quad (3.49)$$

where $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ is a quantum channel, and the last equality is seen simply by convexity of the trace norm.

We first upper bound the trace-norm contraction in terms of the χ^2 contraction, which is a generalization of a result in [Rus94]:

Theorem 34. *For all $\alpha \in (0, 1]$, and a quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$,*

$$\bar{\eta}_{\text{tr}}(T) \leq \sqrt{\bar{\eta}_\chi^\alpha(T)} \quad (3.50)$$

Proof. From Lemma 26, we have that $\|T(\rho - \sigma)\|_1^2 \leq \chi_\alpha^2(T(\rho), T(\sigma))$, for all $\rho, \sigma \in \mathcal{S}_d$. Let N be traceless and hermitian, and note that it can be written as $N = N_+ - N_-$, where N_+, N_- are positive definite and orthogonal in their support. Now let $P = |N| / \|N\|_1$ and recall that $|N| = N_+ + N_-$, then we get $\text{tr}[NP^{-\alpha}NP^{\alpha-1}] = \|N\|_1^2$, for every $\alpha \in (0, 1]$. Also,

$$\frac{\|T(N)\|_1^2}{\|N\|_1^2} \leq \frac{\text{tr}[T(N)T(P)^{-\alpha}T(N)T(P)^{\alpha-1}]}{\text{tr}[NP^{-\alpha}NP^{\alpha-1}]} \quad (3.51)$$

where the inequality is in the numerator, and the denominators are equal, by the previous observation. Taking the supremum over all traceless hermitian N on the left hand side and identifying $\rho - \sigma = N$, $P = \sigma$ then gives desired result. \square

We now provide a lower bound to the trace norm contraction for primitive channels:

Theorem 35. *Given a quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$,*

$$\bar{\eta}_\chi^{\alpha=1/2}(T) \leq \bar{\eta}_{\text{tr}}(T) \quad (3.52)$$

First we introduce an eigenvalue type min-max characterization of the χ^2 -contraction, and then show that this eigenvalue must be smaller than the trace norm-contraction.

Let $P > 0$, and consider the following eigenvalue equation:

$$\hat{\Gamma}|A\rangle \equiv \hat{\Omega}_P^{-1} \hat{\Gamma}^\dagger \hat{\Omega}_{T(P)} \hat{\Gamma}|A\rangle = \lambda|A\rangle, \quad (3.53)$$

where $\Omega_X \equiv \Omega_X^{\alpha=1/2}$. If T has a non-trivial kernel, then $\Omega_{T(P)}$ should be understood in terms of the pseudo-inverse. First note that Γ is a quantum channel, so its spectrum is bounded by one, and that it reaches one for $A = P$. Also note that Γ is similar to a

hermitian operator, so it has all real eigenvalues, so we can take the eigenvectors to be hermitian. Then rewriting Eqn. (3.53) as $\hat{T}^\dagger \hat{\Omega}_{T(P)} \hat{T} |A\rangle = \lambda \hat{\Omega}_P |A\rangle$, we can express the second largest eigenvalue as:

$$\begin{aligned} \lambda_1(T, P) &= \sup_{\langle N | \Omega_P(P) \rangle = 0, N = N^\dagger} \frac{\langle N | \hat{T}^\dagger \hat{\Omega}_{T(P)} \hat{T} | N \rangle}{\langle N | \hat{\Omega}_P | N \rangle} \\ &= \sup_{\text{tr } N = 0, N = N^\dagger} \frac{\text{tr}[T(N)T(P)^{-1/2}T(N)T(P)^{-1/2}]}{\text{tr}[NP^{-1/2}NP^{-1/2}]} . \end{aligned} \quad (3.54)$$

Clearly, by maximizing over all P , one recovers $\bar{\eta}_\chi^{1/2}(T)$. We now prove the above theorem:

Proof. Let N_1 be the eigenvector for which λ_1 satisfies the eigenvalue Eqn. (3.53), and recall that N_1 is Hermitian and traceless. Then,

$$\lambda_1 \|N_1\|_1 = \|\Gamma(T(N_1))\|_1 \leq \|T(N_1)\|_1 \quad (3.55)$$

because Γ is a channel, and

$$\lambda_1 \leq \frac{\|T(N_1)\|_1}{\|N_1\|_1} \leq \sup_{\text{tr } N = 0, N^\dagger = N} \frac{\|T(N)\|_1}{\|N\|_1} = \bar{\eta}_{tr}, \quad (3.56)$$

taking the supremum over positive P completes the proof. \square

Remark: Theorem 35 gives a computable lower bound to the trace norm contraction. A key subtlety in the argument is that $[\Omega_P(A)]^{-1} = \sqrt{P}A\sqrt{P}$ is a completely positive (CP) map (with a single Kraus operator \sqrt{P}) which implies that Γ is a quantum channel. In general, Ω_P is not even positivity preserving. Another exception is the monotone metric associated with the usual logarithmic relative entropy for which $k(w) = \frac{\log w}{w-1}$. It is well-known [AN00, Pet86, LR99] that $\Omega_P^{\log}(A)$ can be written as

$$\Omega_P^{\log}(A) = \int_0^\infty \frac{1}{P+xI} A \frac{1}{P+xI} dx \quad (3.57)$$

which is clearly CP. An analogous lower bound was shown in [LR99] for this map using a similar argument. Clearly, this can be extended to any monotone metric for which Ω_P is CP; however, we do not know of any other examples.

Very little is known about the ordering of the general $\bar{\eta}_k$ contraction coefficients. In particular, We do not know whether $\bar{\eta}_\chi^{\log}$ is smaller or larger than $\bar{\eta}_\chi^{\alpha=1/2}$. However, it is known [LR99] that $\bar{\eta}_k$ are not all identical for different $k \in \mathcal{K}$; because examples can be constructed using non-unital qubit channels. Theorem 34 can readily be extended to any metric associated with $k \in \mathcal{K}$. However, it seems unlikely that Theorem 35 holds in general,. Thus, we can conclude

$$\max\{\bar{\eta}_\chi^{\alpha=1/2}(T), \bar{\eta}_\chi^{\log}(T)\} \leq \bar{\eta}_{tr}(T) \leq \inf_{k \in \mathcal{K}} \sqrt{\bar{\eta}_\chi^k(T)} . \quad (3.58)$$

Note that if instead of maximizing over all P we only consider contraction of the map to the steady state, and denote it $\eta(T) = \bar{\eta}(T)_{P=\sigma}$, then we recover the convergence measure of Eqn. (2.39), and one immediately gets:

$$\eta_\chi^\alpha(T) \leq \eta_{\text{tr}}(T) \leq \bar{\eta}_{\text{tr}}(T) \leq 1 \quad (3.59)$$

Combining this with the previous bounds above, we have

$$\lambda_1 \leq s_1^{\alpha=1/2} = \eta_\chi^{\alpha=1/2} \leq \bar{\eta}_\chi^{\alpha=1/2} \leq \bar{\eta}_{\text{tr}} \leq \sqrt{\bar{\eta}_\chi^{\alpha=1/2}}. \quad (3.60)$$

Moreover, $k(w) = \sqrt{w}$ on the right can be replaced by any $k \in \mathcal{K}$, and that on the left by $k(w) = (w-1)^{-1} \log w$. It is very tempting to conjecture that $\eta_{\text{tr}}^2 \leq \eta_\chi^\alpha$, and/or that $\bar{\eta}_{\text{tr}} \leq \sqrt{\eta_\chi^{\alpha=1/2}}$, but simple numerical counterexamples show these to be false.

3.4 Quantum Detailed Balance

The detailed balance condition is often crucial in the analysis of classical Markov chain mixing times, as it ensures several convenient properties of the Markov chain. In particular, it implies that the classical probability distribution with respect to which the stochastic map is detailed balanced is a fixed point of the chain. Furthermore, detailed balanced stochastic maps have a real spectrum. In this section we generalize the notion of classical detailed balance to quantum Markov chains. Alternative definitions of quantum detailed balance have been given in the literature: [Fri90, MS98, Maj84, Ali76] and references therein. Central to our approach is the operator Q_k as previously introduced in Lemma 30. In the literature for classical Markov chains an analogous matrix exists and is often referred to as the discriminant.

Definition 36. For a quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ and a state $\sigma \in \mathcal{S}_d^+$ with corresponding inversion Ω_σ^k as defined in Eqn. (3.9), we define the quantum discriminant of T as,

$$Q_k = [\Omega_\sigma^k]^{1/2} \circ T \circ [\Omega_\sigma^k]^{-1/2}. \quad (3.61)$$

We recall that the convergence of an arbitrary quantum Markov process can be bounded by the singular values of \hat{Q}_k . Classical detailed balanced Markov chains have the property that the corresponding discriminant becomes symmetric. We shall therefore define the quantum generalization by requiring that for a quantum detailed balanced process

$$Q_k^* = Q_k. \quad (3.62)$$

This immediately allows to make a statement about the spectrum of quantum detailed balanced maps. Due to the hermiticity of the matrix representation of the map Eqn. (3.61) we can immediately deduce, just as for classically case, that the quantum channel T has a real spectrum. For detailed balanced maps, the second largest eigenvalue in magnitude coincides with the second largest singular value. Furthermore, we would like to point out that this is actually not just a single condition for quantum detailed balance but a whole family. Hence every different inversion Ω_σ^k gives rise to a different condition for detailed balance. We therefore define as the quantum generalization of detailed balance:

Definition 37. For a quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ and a state $\sigma \in \mathcal{S}_d^+$, we say that T obeys k -detailed balance with respect to σ with $k \in \mathcal{K}$, when

$$[\Omega_\sigma^k]^{-1} \circ T^* = T \circ [\Omega_\sigma^k]^{-1}. \quad (3.63)$$

A consequence of this definition is that σ is a fixed point of T .

Lemma 38. Let $\sigma \in \mathcal{S}_d$ be a state and T a quantum channel that satisfies the detailed balance condition of Eqn. 37 with respect to Ω_σ^k , then σ is a steady state of T .

Proof. Recall that the inverse is given by $[\Omega_\sigma^k]^{-1} = R_\sigma f(\Delta_{\sigma,\sigma})$, where $f(w) = 1/k(w)$. Hence, since $k(1) = f(1) = 1$, we have

$$[\Omega_\sigma^k]^{-1}(\mathbb{1}) = R_\sigma f(\Delta_{\sigma,\sigma})\mathbb{1} = R_\sigma \mathbb{1} = \sigma. \quad (3.64)$$

Now, since furthermore $T^*(\mathbb{1}) = \mathbb{1}$, we have that

$$T(\sigma) = T \circ [\Omega_\sigma^k]^{-1}(\mathbb{1}) = [\Omega_\sigma^k]^{-1} \circ T^*(\mathbb{1}) = [\Omega_\sigma^k]^{-1}(\mathbb{1}) = \sigma. \quad (3.65)$$

□

Given a probability distribution on some set of states, it is desirable to have a simple criterium to check whether a completely positive map obeys detailed balance with respect to the state generated from the distribution. This criterium may then serve to set up a Markov chain that actually converges to the desired steady state.

Proposition 39. Let $\{|i\rangle\}_i$ be a complete orthonormal basis of \mathcal{H} and let $\{\mu_i\}_i$ be a probability distribution on this basis. Furthermore, assume that a quantum channel T obeys

$$\frac{\mu_n}{k(\mu_m/\mu_n)} \langle i | T(|n\rangle\langle m|) | j \rangle = \frac{\mu_i}{k(\mu_j/\mu_i)} \langle m | T(|j\rangle\langle i|) | n \rangle, \quad (3.66)$$

then $\sigma = \sum_i \mu_i |i\rangle\langle i|$ and T obey the detailed balance condition with respect to Ω_σ^k .

Proof. Note that $\{|i\rangle\langle j|\}_{ij}$ forms a complete and orthonormal basis in the space \mathcal{M}_d with respect to the Hilbert-Schmidt scalar product. We can therefore express Eqn. (3.63) in this basis. The individual entries are equal due to

$$\begin{aligned} \text{tr} \left[(|m\rangle\langle n|)^\dagger [\Omega_\sigma^k]^{-1} \circ T^*(|j\rangle\langle i|) \right] &= \mu_n k^{-1} (\mu_m/\mu_n) \text{tr} \left[T(|m\rangle\langle n|)^\dagger (|j\rangle\langle i|) \right] \\ &= \mu_n k^{-1} (\mu_m/\mu_n) \langle i | T(|n\rangle\langle m|) | j \rangle = \mu_i k^{-1} (\mu_j/\mu_i) \langle m | T(|j\rangle\langle i|) | n \rangle = \\ &= \mu_i k^{-1} (\mu_j/\mu_i) \text{tr} \left[(|m\rangle\langle n|)^\dagger T(|j\rangle\langle i|) \right] = \text{tr} \left[(|m\rangle\langle n|)^\dagger T \circ [\Omega_\sigma^k]^{-1} (|j\rangle\langle i|) \right]. \end{aligned} \quad (3.67)$$

□

Remark: We note that the different quantum detailed balance conditions coincide for classical channels, i.e. for stochastic processes that are included in the framework of quantum channels. Define the following "classical" Kraus operators:

$$A_{ij}^{cl} = \sqrt{P_{ij}} |i\rangle\langle j| \quad \text{and a state, } \sigma = \sum_i \mu_i |i\rangle\langle i|. \quad (3.68)$$

In this case, the condition of Proposition 39 reduces to the classical condition. This can be seen when considering the channel $T^{cl}(\rho) = \sum_{ij} A_{ij}^{cl} \rho A_{ij}^{cl\dagger}$ and checking for detailed balance with respect to sigma, since

$$\frac{\mu_m}{k(\mu_n/\mu_m)} \langle i | T^{cl}(|n\rangle\langle m|) | j \rangle = \frac{\mu_m}{k(\mu_n/\mu_m)} \delta_{nm} \delta_{ij} P_{in}$$

and

$$\frac{\mu_i}{k(\mu_j/\mu_i)} \langle i | T^{cl}(|n\rangle\langle m|) | j \rangle = \frac{\mu_i}{k(\mu_j/\mu_i)} \delta_{nm} \delta_{ij} P_{ni}. \quad (3.69)$$

However since $k(1) = 1$ we are just left with the classical detailed balance condition $\mu_i P_{ni} = \mu_n P_{in}$ for all pairs i, n .

A natural question to ask is therefore, whether the different detailed balance condition are all identical. To see that this is not the case, consider the example given by the Kraus operators of a single qubit, i.e. $\mathcal{H} = \mathbb{C}^2$,

$$A_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad A_2 = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}. \quad (3.70)$$

This channel has the unique fixed point

$$\sigma = \frac{1}{6} \begin{pmatrix} 5 & 1 \\ 1 & 1 \end{pmatrix}. \quad (3.71)$$

From this channel it is now possible to construct a channel that obeys detailed balance with respect to the inversion given by choosing $k(w) = w^{-1/2}$, that is the inversion reads $\Omega_\sigma^{\alpha=1/2} = L_\sigma^{-1/2} R_\sigma^{-1/2}$. We consider therefore the symmetrized map,

$$T_s = \left[\Omega_\sigma^{\alpha=1/2} \right]^{-1} \circ T^* \circ \Omega_\sigma^{\alpha=1/2} \circ T. \quad (3.72)$$

For the specific instance where $\Omega_\sigma^{\alpha=1/2}$ is given as above, we are assured that the map T_s is again a quantum channel, because one immediately finds the Kraus representation for $T_s(\rho) = \sum_{ij} B_{ij} \rho B_{ij}^\dagger$ as $B_{ij} = \sqrt{\sigma} A_i^\dagger [\sqrt{\sigma}]^{-1} A_j$. The individual Kraus operators read,

$$\begin{aligned} B_{11} &= \frac{3}{5} \begin{pmatrix} 1 & 1 \\ 1/2 & 1/2 \end{pmatrix} \quad \text{and} \quad B_{12} = \frac{\sqrt{2}}{5} \begin{pmatrix} 1 & -1 \\ 1/2 & -1/2 \end{pmatrix}, \\ B_{21} &= \frac{\sqrt{2}}{20} \begin{pmatrix} 3 & 3 \\ -1 & -1 \end{pmatrix} \quad \text{and} \quad B_{22} = \frac{1}{5} \begin{pmatrix} 3 & -3 \\ -1 & 1 \end{pmatrix}. \end{aligned} \quad (3.73)$$

The channel T_s satisfies detailed balance with respect to $\Omega_\sigma^{\alpha=1/2}$ by construction. This channel however does not satisfy detailed balance with respect to the inversion $\Omega_\sigma^{Bures} = 2[L_\sigma + R_\sigma]^{-1}$ as can be seen directly by evaluating the detailed balance condition in terms of the matrix representations,

$$\left[\hat{\Omega}_\sigma^{Bures} \right]^{-1} \cdot \hat{T}_s^\dagger - \hat{T}_s \cdot \left[\hat{\Omega}_\sigma^{Bures} \right]^{-1} = \frac{7}{600} [\mathbb{1} \otimes Y + Y \otimes \mathbb{1}], \quad (3.74)$$

where

$$Y = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \quad (3.75)$$

The family of quantum detailed balance conditions is therefore much richer than the classical counterpart.

3.5 Quantum Cheeger's Inequality

In the context of classical stochastic processes a very powerful formalism has been developed, often referred to as the conductance bound or Cheeger's inequality, to bound convergence rates of stochastic processes. We will generalize this to the quantum setting in this section. Similar results have appeared in [Has07]. The gap of the map S_k is defined as the difference between the largest and second largest eigenvalue, $\Delta = 1 - \lambda_1$. The gap can be characterized in a variational fashion [HJ07].

Proposition 40. *The gap of the map $S_k = [\Omega_\sigma^k]^{-1/2} \circ T^* \circ \Omega_\sigma^k \circ T \circ [\Omega_\sigma^k]^{-1/2}$ is given by*

$$\Delta = \min_{X \in M_d} \frac{\langle X, (id - S_k)X \rangle}{\frac{1}{2} \|(X \otimes \sqrt{\sigma} - \sqrt{\sigma} \otimes X)\|_{HS}^2}, \quad (3.76)$$

where $\|A\|_{HS}^2 = \text{tr}[A^\dagger A]$ denotes the standard Hilbert-Schmidt norm and $\langle \cdot, \cdot \rangle$ the corresponding Hilbert-Schmidt scalar product.

Proof. The eigenvector that corresponds to the eigenvalue $\lambda_0 = 1$ of S_k is given by $\sqrt{\sigma}$. The gap can therefore be written as [HJ07]:

$$\begin{aligned} \Delta &= \min_{X \in M_d; \text{tr}[X\sqrt{\sigma}] = 0} 1 - \frac{\text{tr}[X^\dagger S(X)]}{\text{tr}[X^\dagger X]} \\ &= \min_{X \in M_d; \text{tr}[X\sqrt{\sigma}] = 0} \frac{\text{tr}[X^\dagger (X - S(X))]}{\text{tr}[X^\dagger X] - \text{tr}[X\sqrt{\sigma}]^2} \\ &= \min_{X \in M_d} \frac{\text{tr}[X^\dagger (X - S(X))]}{\frac{1}{2} \|(X \otimes \sqrt{\sigma} - \sqrt{\sigma} \otimes X)\|_{HS}^2}, \end{aligned} \quad (3.77)$$

Note that the constrained $\text{tr}[X\sqrt{\sigma}] = 0$ can be dropped in the last line. Suppose that $\text{tr}[X\sqrt{\sigma}] = c$, we can then define $X' = X - c\sqrt{\sigma}$ and vary X' since the equation is invariant under such shifts. \square

Throughout the remainder of this section we consider unital quantum channels, i.e. maps which obey $T(\mathbb{1}) = \mathbb{1}$. For this case it is ensured that already the simple map $S = T^* \circ T$ has a spectrum that is contained in $[0, 1]$, since all Ω_σ^k coincide and correspond to the identity map. The χ^2 -divergence just reduces to the standard Hilbert-Schmidt inner product times a prefactor given by the dimension of the space d . In the case of a detailed balanced stochastic map it even suffices to just consider the map itself. In either case we will denote the corresponding map as S from now on. The variational characterization of the gap Δ now allows us to give an upper as well as a lower bound to the second largest eigenvalue of S .

Lemma 41. *Let $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a unital quantum channel. Then the second largest eigenvalue λ_1 of its symmetrization $S = T^* \circ T$, is bounded by,*

$$1 - 2h \leq \lambda_1 \leq 1 - \frac{1}{2}h^2, \quad (3.78)$$

where h is Cheeger's constant defined as,

$$h = \min_{\Pi_A, \text{tr}[\Pi_A] \leq d/2} \frac{\text{tr}[(\mathbb{1} - \Pi_A) S(\Pi_A)]}{\text{tr}[\Pi_A]}. \quad (3.79)$$

The minimum is to be taken over all projectors Π_A on the space $A \subset \mathcal{M}_d$, so that $\text{tr}[\Pi_A] \leq d/2$.

Proof. An upper bound to the gap is immediately found by choosing $X = \Pi_A$. Due to Proposition (40) we can write:

$$\begin{aligned} \Delta &\leq \frac{\text{tr}[\Pi_A(\text{id} - S)(\Pi_A)]}{\text{tr}[\Pi_A^2] - \frac{1}{d} \text{tr}[\Pi_A]^2} \\ &= \frac{\text{tr}[(\mathbb{1} - \Pi_A)S(\Pi_A)]}{\frac{1}{d} \text{tr}[(\mathbb{1} - \Pi_A)] \text{tr}[\Pi_A]} \leq 2h, \end{aligned} \quad (3.80)$$

where in the last line we have used that $\text{tr}[\mathbb{1} - \Pi_A] \geq d/2$.

For the lower bound, we can restrict the minimization in Eqn. (3.79) to diagonal projections. Furthermore, when considering only unital quantum channels, it is possible to reduce the problem of bounding the gap Δ to that of bounding the gap of a classical stochastic process. To see this, let us work in the basis where the eigenvector $X_1 \in \mathcal{M}_d$ corresponding to λ_1 is diagonal. We shall assume wlog that $X_1^\dagger = X_1$. In this basis, we can write $X = \sum x_i |i\rangle\langle i|$. The numerator then becomes

$$\begin{aligned} \text{tr}[X^\dagger(X - S(X))] &= \sum_{ij} x_i x_j (\text{tr}[|i\rangle\langle i||j\rangle\langle j|] - \text{tr}[|i\rangle\langle i|S(|j\rangle\langle j|)]) \\ &= \sum_i x_i^2 - \sum_{ij} x_i x_j P_{ij} = \frac{1}{2} \sum_{ij} P_{ij} (x_i - x_j)^2. \end{aligned} \quad (3.81)$$

We introduced the matrix $P_{ij} = \langle i|S(|j\rangle\langle j|)|i\rangle$, which is a symmetric non-negative matrix which obeys $P_{ij} \geq 0$, $\sum_i P_{ij} = 1$ and $P^T = P$. Hence P is doubly stochastic. Performing the same reduction in the denominator we obtain

$$\frac{1}{2d} \|(X \otimes \mathbb{1} - \mathbb{1} \otimes X)\|_{HS}^2 = \frac{1}{2d} \sum_{ij} (x_i - x_j)^2 \quad (3.82)$$

Hence, we arrive at the classical version of Mihail's Identity [Mih89],

$$\Delta = \min_{\{x_i\}} \frac{\sum_{ij} P_{ij} (x_i - x_j)^2}{1/d \sum_{ij} (x_i - x_j)^2}. \quad (3.83)$$

Given the classical version of Mihail's identity, the proof of the lower bound is the same as in the classical case. For completeness we repeat it here.

First, we define, $z_i \equiv |x_i|x_i$ and write,

$$\begin{aligned} \sum_{ij} P_{ij} |z_i - z_j| &= \sum_{ij} P_{ij} ||x_i|x_i - |x_j|x_j|| \leq \sum_{ij} \sqrt{P_{ij}} \sqrt{P_{ij}} (|x_i| + |x_j|) (x_i - x_j) \\ &\leq \sqrt{\sum_{ij} P_{ij} (x_i - x_j)^2} \sqrt{\sum_{ij} P_{ij} (|x_i| + |x_j|)^2}, \end{aligned} \quad (3.84)$$

where we used Cauchy-Schwartz in the last step. Consider now,

$$\sum_{ij} P_{ij} (|x_i| + |x_j|)^2 = 2(\sum_i x_i^2 + \sum_{ij} |x_i| P_{ij} |x_j|) \leq 4 \sum_i |x_i|^2. \quad (3.85)$$

Furthermore, note that we can bound,

$$1/d \sum_{ij} (x_i - x_j)^2 \leq 2/d \sum_{ij} x_i^2 = 2 \sum_i |z_i|. \quad (3.86)$$

We are therefore left with a lower bound to Mihail's identity, which holds for all choices of $\{x_i\}$

$$\frac{1}{2} \left(\frac{\sum_{ij} P_{ij} |z_i - z_j|}{2 \sum_i |z_i|} \right)^2 \leq \frac{\sum_{ij} P_{ij} (x_i - x_j)^2}{1/d \sum_{ij} (x_i - x_j)^2}. \quad (3.87)$$

We shall now assume, that $x_i \geq 0$ everywhere and we can hence drop the absolute values in the definition for the z_i . This assumption is valid since we are free in adding an arbitrary constant $x_i \rightarrow x_i + c$ to make all x_i positive. Note that we therefore are left with a lower bound to the gap of the form,

$$\Delta \geq \frac{1}{2} \left(\frac{\sum_{ij} P_{ij} |x_i^2 - x_j^2|}{2 \sum_i x_i^2} \right)^2 \quad (3.88)$$

Let's focus on the right side of the inequality. Since,

$$2 \sum_{ij: x_i \geq x_j} P_{ij} (x_i^2 - x_j^2) = 4 \sum_{ij: x_i \geq x_j} P_{ij} \int_{x_j}^{x_i} t \, dt = 4 \int_0^\infty t \sum_{ij: x_i > t \geq x_j} P_{ij} \, dt, \quad (3.89)$$

and furthermore,

$$\sum_{ij: x_i > t \geq x_j} P_{ij} = \sum_{i \in A(t)} \sum_{j \in A^c(t)} P_{ij} \quad \text{where, } A(t) \equiv \{i | x_i \geq t\}, \quad (3.90)$$

we can bound,

$$4 \int_0^\infty t \sum_{ij: x_i > t \geq x_j} P_{ij} \, dt \geq h 4 \int_0^\infty t \sum_{i \in A(t)} \Theta(t - x_i) \, dt = 2 h \left(\sum_i x_i^2 \right), \quad (3.91)$$

where we defined h as in the same fashion as above. We have therefore found the desired lower bound for the spectral gap of the map S . \square

3.5.1 Example: Conductance bound for unital qubit channels

A convenient basis for the matrix space \mathcal{M}_2 associated with the Hilbert space $\mathcal{H} = \mathbb{C}^2$ is given in terms of the Pauli basis $\{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}$. In this basis a density matrix $\rho \in \mathcal{S}_2$ can be parametrized in terms of its Bloch vector $\mathbf{r} \in \mathbb{R}^3$. In the Bloch representation the density matrix reads $\rho = \frac{1}{2}(\mathbb{1} + \mathbf{r} \cdot \Sigma)$, where $\Sigma = (\sigma_x, \sigma_y, \sigma_z)$. It is also straight forward to determine the matrix representation of a quantum channel $T : \mathcal{M}_2 \rightarrow \mathcal{M}_2$ with respect to the Pauli basis. A general channel can be written as a matrix $\hat{T} \in \mathcal{M}_4$.

$$\hat{T} = \begin{pmatrix} 1 & 0 \\ \mathbf{t} & \mathbf{L} \end{pmatrix}. \quad (3.92)$$

The channel acts on a density matrix via $T(\rho) = T(\frac{1}{2}(\mathbb{1} + \mathbf{r} \cdot \Sigma)) = \frac{1}{2}(\mathbb{1} + (\mathbf{t} + \mathbf{L}\mathbf{r}) \cdot \Sigma)$. It can be shown, that the map T is unital if and only if $\mathbf{t} = 0$. Let us now consider the

optimization for Cheeger's constant h as given in Lemma (41). Given the constraint, we have to vary all one dimensional projectors $\Pi_A = |\psi\rangle\langle\psi|$ with $\|\psi\|_2 = 1$, so that

$$h = \min_{|\psi\rangle \in \mathbb{C}^2} \text{tr} [(\mathbb{1} - |\psi\rangle\langle\psi|) S(|\psi\rangle\langle\psi|)]. \quad (3.93)$$

The symmetrized map S of the unital channel T , with $\mathbf{t} = 0$, now assumes the matrix representation,

$$\hat{S} = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{L}^\dagger \mathbf{L} \end{pmatrix}. \quad (3.94)$$

Furthermore note, that any projector $|\psi\rangle\langle\psi| \in \mathcal{S}_2$ can be parametrized via a Bloch vector $\mathbf{a} \in \mathbb{R}^3$ that obeys $\|\mathbf{a}\|_2 = 1$. The minimization for Cheeger's constant reduces therefore to

$$h = \min_{\|\mathbf{a}\|_2=1} 1 - \langle \mathbf{a} | \mathbf{L}^\dagger \mathbf{L} | \mathbf{a} \rangle, \quad (3.95)$$

where $\langle \mathbf{a} | \mathbf{b} \rangle$ denotes the canonical scalar product in \mathbb{R}^3 . The minimum is attained when \mathbf{a} is the eigenvector associated with the largest eigenvalue s_1^2 of the matrix $\mathbf{L}^\dagger \mathbf{L}$. Hence for an arbitrary single qubit unital channel, Cheeger's constant is given by $h = 1 - s_1^2$, where s_1 is the largest singular value of the matrix \mathbf{L} and hence the second largest singular value of the channel T . We see that the conductance bound as stated in (Lemma 41) is indeed satisfied, since

$$2s_1^2 - 1 \leq s_1^2 \leq \frac{1}{2}(1 + s_1^2). \quad (3.96)$$

The Cutoff Phenomenon

Synopsis:

We derive upper and lower bounds on the convergence behavior of certain classes of one-parameter quantum dynamical semigroups. The classes we consider consist of tensor product channels and of channels with commuting Liouvillians. We introduce the notion of *Cutoff Phenomenon* in the setting of quantum information theory, and show how it exemplifies the fact that the convergence of (quantum) stochastic processes is not solely governed by the spectral gap of the transition map. We identify a number of situations withing the setting of product channels where a strict cutoff can be proved.

The outline of the chapter is as follows. In Section 4.1, we give a definition of the Cutoff Phenomenon in the quantum setting. In Section 4.2, we state and prove the main results, namely cutoff-type bounds for time-evolutions due to commuting Liouvillians or tensor product channels. Section 4.3 illustrates these main results by various examples.

Based on:

A Cutoff Phenomenon for Quantum Markov Chains

M. J. Kastoryano, D. Reeb, M. M. Wolf

arXiv:1111.2123, (submitted to J. Phys. A)

4.1 The Cutoff Phenomenon

Often times, the relevant question when considering the convergence behavior of open systems is “how long does the process have to run before it reaches equilibrium?” To make precise statements about convergence, it is usually necessary to consider how the time to convergence scales with the system size. For instance, one would like to know how fast, as a function of the lattice size, a given dynamical process on a lattice converges to its steady state.

It was observed a while ago in the setting of classical Markov chains, that for a special set of chains this question can be answered exactly when the size of the system becomes large. This behavior, which has been coined the *Cutoff Phenomenon* [Dia96, BD92, SC04], characterizes the situation when for some (possibly long) period of time some information from the initial state is perfectly preserved until a critical time. Shortly after this critical time, however, essentially no information of the initial state can be recovered from the time-evolved state anymore. For large system sizes n , the convergence of the channel as a function of time t will look like a step function at the *cutoff time* t_n (see Fig. 4.1a).

This behavior has been observed and proved to occur in a number of interesting examples of classical Markov chains. One case where this phenomenon is particularly pronounced, and which triggered widespread popular interest, is in card shuffling, where it was shown that a deck of n cards is well mixed after exactly $3/2 \log n$ riffle shuffles, and poorly mixed under $3/2 \log n$ riffle shuffles (when n becomes large) [BD92]. In particular, this guarantees casino owners that if their dealers riffle shuffle their 52 poker cards seven or more times before each draw, then they do not need to worry about about players trying to improve their odds by counting cards¹. Several other processes have also been shown to exhibit cutoffs, including random walks on graphs with a discrete group structure, birth and death type chains, and some Monte Carlo sampling methods [SC04, DLP08, LS09]. We now give a formal definition in the quantum setting:

Definition 42 (Cutoff). *Let $T_t^{(n)}$ be a sequence, indexed by the “system size” n , of one-parameter semigroups of quantum channels. We say that $T_t^{(n)}$ exhibits a cutoff (in trace-norm) at times t_n , if for any real $c > 0$:*

$$\begin{aligned} c < 1 & \Rightarrow \lim_{n \rightarrow \infty} \eta_{\text{tr}}[T_{ct_n}^{(n)}] = 1, \\ c > 1 & \Rightarrow \lim_{n \rightarrow \infty} \eta_{\text{tr}}[T_{ct_n}^{(n)}] = 0. \end{aligned}$$

We point out that this is not the only definition of a cutoff, and in a sense it is an incomplete one, as it does not provide detailed information about the cutoff window, i.e. the width of the drop-off at t_n (see Fig. 4.1a).

In many situations, it is difficult to prove an actual cutoff, whereas it might be easier to show a weaker statement which gives only the precise order of magnitude of the time to convergence:

¹The punchline in this example is that the cutoff behavior can be exploited to perform a spectacular magic trick called “Premo” [Wil12], where the magician can guess a card chosen randomly by a person from the audience who afterwards shuffles the deck a few times. The information in the initial configuration (i.e., the identity of the card that the person put on top) is preserved and can be identified before a “time” of exactly $\frac{3}{2} \log n$ riffle shuffles (for large n).

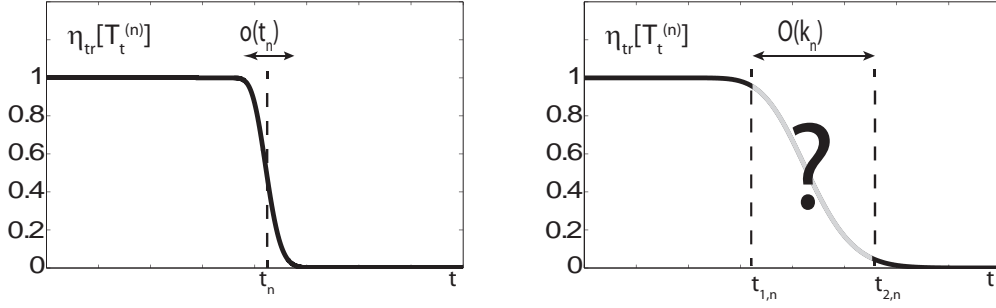


FIGURE 4.1: Behavior of the convergence measure $\eta_{\text{tr}}[T_t^{(n)}]$ as a function of evolution time t for a one-parameter semigroup of channels that exhibits a cutoff (for large system size n). At time $t = t_n$, essentially all dependence on the initial state vanishes in a small window of time $o(t_n)$. In the case of pre-cutoff, the convergence measure will be close to 1 for times $t < t_{1,n} = \Theta(k_n)$ and close to 0 for $t > t_{2,n} = \Theta(k_n)$, but now the gap between $t_{1,n}$ and $t_{2,n}$ may be of order $O(k_n)$.

Definition 43 (Pre-cutoff). *With the same assumptions as in Definition 42, we say that $T_t^{(n)}$ exhibits a pre-cutoff (in trace-norm) of order $\Theta(k_n)$ for some sequence k_n , if there exist times $t_{1,n}$ and $t_{2,n}$ that are both of order $\Theta(k_n)$, such that*

$$\begin{aligned} c < 1 &\quad \Rightarrow \quad \lim_{n \rightarrow \infty} \eta_{\text{tr}}[T_{ct_{1,n}}^{(n)}] = 1, \\ c > 1 &\quad \Rightarrow \quad \lim_{n \rightarrow \infty} \eta_{\text{tr}}[T_{ct_{2,n}}^{(n)}] = 0. \end{aligned}$$

We again emphasize that the cutoff phenomenon does not cover all types of pre-asymptotic behavior. For example, a process can follow a polynomial decay for a certain amount of time before it settles into the asymptotic (exponential) regime. We also point out that the Cutoff Phenomenon can be defined with respect to any monotone distance measure (even unbounded ones like the χ^2 -divergence). A cutoff in one distance measure does not imply a cutoff in another measure. This suggests that obtaining such a tight contraction estimate might in some cases reveal information about only one specific facet of the convergence behavior, associated with the given distance measure. For instance, the trace-norm convergence measure of a channel tells us whether a single bit of classical information is preserved after a certain time, but makes no statement about the entire amount of information or the preservation of quantum information.

4.2 Main Results

In this section, we present two situations which exhibit behavior related to the Cutoff Phenomenon introduced above: commuting Liouvillians and, as a more restrictive but still relevant class, Liouvillians acting independently on subsystems. In the latter situation, the convergence behavior of the constituent channels is known, and we ask how the convergence of the tensor product of channels behaves. We show in general terms that a sequence of tensor product channels exhibits a pre-cutoff of order $\Theta(\log n)$, where n is the number of tensor factors. In the next section, we discuss some specific situations of this kind where an actual cutoff occurs, which includes the dissipative preparation of stabilizer states.

Our first theorem provides a general upper bound on the convergence measure of a channel from a one-parameter semigroup whose Liouvillian is composed of commuting parts, i.e. $\mathcal{L} = \sum_j \mathcal{L}_j$ where $[\mathcal{L}_j, \mathcal{L}_k] = 0$. In this case, the gap of the full Liouvillian is at least the minimum of the gaps of its constituent parts. We show that in this context the time to convergence is upper bounded by $O(\log n)$ times the convergence time of the “slowest” constituent channel, where n is the number of commuting terms in the Liouvillian.

Theorem 44 (Convergence for commuting Liouvillians). *Let $\mathcal{L}_j : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be Liouvillians which commute, i.e. $[\mathcal{L}_j, \mathcal{L}_k] = 0$ for all $j, k = 1, \dots, n$. Define $\mathcal{L} \equiv \sum_j \mathcal{L}_j$, and the corresponding semigroups of channels $T_{t,j} \equiv e^{t\mathcal{L}_j}$ and $T_t \equiv e^{t\mathcal{L}}$ ($t \geq 0$). Then:*

$$\eta_{\text{tr}}[T_t] \leq \sum_j \eta_{\text{tr}}[T_{t,j}]. \quad (4.1)$$

Proof. The theorem is proved by induction. Let $T_{\varphi,1}$ be the projector onto the asymptotic space of $T_{t,1}$, and let $T_{\varphi,j \neq 1}$ be the projector onto the asymptotic space of $T_{t,j \neq 1} \equiv e^{t\sum_{j \neq 1} \mathcal{L}_j}$, see Eqn. (2.19). Note that $T_t = T_{t,1}T_{t,j \neq 1}$ and $T_\varphi = T_{\varphi,1}T_{\varphi,j \neq 1}$, by commutativity of the Liouvillians. Then,

$$\begin{aligned} \eta_{\text{tr}}[T_t] &= \frac{1}{2} \sup_{\rho \in \mathcal{S}_d} \|(T_{t,1}T_{t,j \neq 1} - T_{\varphi,1}T_{\varphi,j \neq 1})(\rho)\|_1 \\ &= \frac{1}{2} \sup_{\rho \in \mathcal{S}_d} \|T_{t,1}(T_{t,j \neq 1} - T_{\varphi,j \neq 1})(\rho) + (T_{t,1} - T_{\varphi,1})(T_{\varphi,j \neq 1})(\rho)\|_1 \\ &\leq \eta_{\text{tr}}[T_{t,1}] + \eta_{\text{tr}}[T_{t,j \neq 1}], \end{aligned} \quad (4.2)$$

where the last inequality follows from the triangle inequality, from monotonicity of the trace-norm under quantum channels, and by definition of $\eta_{\text{tr}}[T_t]$. By induction, we get $\eta_{\text{tr}}[T_t] \leq \sum_j \eta_{\text{tr}}[T_{t,j}]$. \square

An immediate consequence of Theorem 44, also using Theorem 20, is that for a system described by n commuting Liouvillians with bounded gaps, the convergence time will be upper bounded by $O(\log n)$. Note however, that commuting Liouvillians should not be confused with classical processes. Indeed, as described in Section 5.1.1, graph states can be prepared dissipatively as stationary states of commuting Liouvillians, whereas these states can be highly entangled (e.g. cluster state).

Our second result gives the general convergence behavior of a tensor power of a one-parameter semigroup of quantum channels. This is a special case of commuting Liouvillians, but where strict upper and lower bounds can be derived as the number of tensor factors becomes large, thereby establishing a pre-cutoff as defined in Section 4.1.

Theorem 45 (Pre-cutoff for tensor powers). *Let $\mathcal{L} : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a Liouvillian with gap $\bar{\lambda}$, and let $T_t \equiv e^{t\mathcal{L}}$ ($t \geq 0$). The sequence of one-parameter semigroups $T_t^{(n)} \equiv T_t^{\otimes n}$ exhibits a pre-cutoff in trace-norm at times $t_{1,n} = \log(n)/2\bar{\lambda}$ and $t_{2,n} = \log(n)/\bar{\lambda}$.*

Proof. Here and below we use the fact that $(T \otimes S)_\varphi = T_\varphi \otimes S_\varphi$ for any pair of quantum channels T and S . To prove the lower bound, let $c \in (0, 1)$:

$$\begin{aligned}
\eta_{\text{tr}}[T_{ct_{1,n}}^{(n)}] &= \sup_{\rho \in \mathcal{S}_{d^n}} d_{\text{tr}}(T_{ct_{1,n}}^{(n)}(\rho), T_\varphi^{(n)}(\rho)) \geq \sup_{\sigma \in \mathcal{S}_d} d_{\text{tr}}((T_{ct_{1,n}}(\sigma))^{\otimes n}, (T_\varphi(\sigma))^{\otimes n}) \\
&\geq 1 - \exp \left[-\frac{1}{2} n \sup_{\sigma \in \mathcal{S}_d} d_{\text{tr}}^2(T_{ct_{1,n}}(\sigma), T_\varphi(\sigma)) \right] \\
&\geq 1 - \exp \left[-\frac{L^2}{2} n e^{-2ct_{1,n}\bar{\lambda}} \right] \\
&= 1 - \exp \left[-\frac{L^2}{2} n^{1-c} \right] \rightarrow 1 \quad (n \rightarrow \infty).
\end{aligned} \tag{4.3}$$

The first inequality is obtained by restricting the supremum to product states $\rho = \sigma^{\otimes n}$, the next from Lemma 46 (see below), and the last follows from Theorem 20 (with some constant $L > 0$). Hence, $\lim_{n \rightarrow \infty} \eta_{\text{tr}}[T_{ct_{1,n}}^{(n)}] = 1$, for $c \in (0, 1)$.

For the upper bound, we apply Theorem 44 to get $\eta_{\text{tr}}[T_t^{(n)}] \leq n\eta_{\text{tr}}[T_t \otimes \text{id}_{n-1}]$, where id_{n-1} is the identity channel on $n-1$ sites. In the following paragraph we show $\eta_{\text{tr}}[T_t \otimes \text{id}_{n-1}] \leq 4d\eta_{\text{tr}}[T_t]$. Now, for any given $c > 1$ one can choose $\nu < \bar{\lambda}$ such that $c\nu/\bar{\lambda} > 1$, and by Theorem 20 one can find R such that $\eta_{\text{tr}}[T_t] \leq Re^{-\nu t}$ for all $t \geq 0$. Combining all this, we finally get that for any $c > 1$,

$$\eta_{\text{tr}}[T_{ct_{2,n}}^{(n)}] \leq 4dn Re^{-\nu ct_{2,n}} = 4Rdn^{1-c\nu/\bar{\lambda}} \rightarrow 0 \quad (n \rightarrow \infty). \tag{4.4}$$

It remains to show that $\eta_{\text{tr}}[T \otimes \text{id}] \leq 4d\eta_{\text{tr}}[T]$ for any channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ and any identity channel $\text{id} : \mathcal{M}_{d'} \rightarrow \mathcal{M}_{d'}$. The inequality (used below) between the norm $\|\cdot\|_{1-1}$ on superoperators induced by the trace-norm and its stabilized version $\|\cdot\|_{cb}$ is proven in [Pau03] (exercise 3.11). In the following, $X \in \mathcal{M}_d \otimes \mathcal{M}_{d'}$ and $A + iB \in \mathcal{M}_d$ denote arbitrary matrices, A and B are Hermitian, and $P, Q \in \mathcal{M}_d$ are positive semidefinite with $PQ = 0$. Further note that $\|A\|_1 = \|(A + iB) + (A - iB)\|_1/2 \leq (\|A + iB\|_1 + \|A - iB\|_1)/2 = \|A + iB\|_1$, and similarly $\|B\|_1 \leq \|A + iB\|_1$. Thus:

$$\begin{aligned}
\eta_{\text{tr}}[T \otimes \text{id}] &= \sup_{\rho \in \mathcal{S}_{dd'}} d_{\text{tr}}(T \otimes \text{id}(\rho), T_\varphi \otimes \text{id}(\rho)) \leq \sup_{\|X\|_1 \leq 1} \frac{1}{2} \|((T - T_\varphi) \otimes \text{id})(X)\|_1 \\
&\leq \frac{1}{2} \|T - T_\varphi\|_{cb} \leq \frac{d}{2} \|T - T_\varphi\|_{1-1} = \frac{d}{2} \sup_{\|A+iB\|_1 \leq 1} \|(T - T_\varphi)(A + iB)\|_1 \\
&\leq d \sup_{\|A\|_1 \leq 1} \|(T - T_\varphi)(A)\|_1 = d \sup_{\|P-Q\|_1 \leq 1} \|(T - T_\varphi)(P - Q)\|_1 \\
&\leq 2d \sup_{\|P\|_1 \leq 1} \|(T - T_\varphi)(P)\|_1 = 4d \eta_{\text{tr}}[T].
\end{aligned}$$

□

Theorem 45 can be generalized to certain cases where $T_t^{(n)}$ is the tensor product of a set of one-parameter semigroups T_t^i that are not all identical. See [BLY06] for the analogous classical result.

Theorem 45 establishes pre-cutoff rather than actual cutoff behavior. But at the end of subsection 4.3.3 we show examples where, for any chosen $r \in [1, 2]$, a cutoff occurs at

times $t_n = \log(n)/r\bar{\lambda}$. This means that $t_{1,n}$ and $t_{2,n}$ in Theorem 45, viewed as upper and lower bounds on the contraction, are tight when expressed in terms of the gap $\bar{\lambda}$.

The following Lemma completes the proof of Theorem 45 (cf. Eqn. (4.3)); the inequalities (4.5) for the Bures distance will be used to show cutoff in Proposition 47. For collections $\rho_i, \sigma_i \in \mathcal{S}_d$ of density matrices ($i = 1, \dots, n$), define $\rho^{(n)} \equiv \bigotimes_{i=1}^n \rho_i$ and similarly $\sigma^{(n)}$.

Lemma 46 (Distances between tensor product states). *Let $\rho_i, \sigma_i \in \mathcal{S}_d$, $i = 1, \dots, n$, and denote by d_{tr}, d_B the trace and Bures distances, respectively. Then the following inequalities hold:*

$$1 - \exp \left[- \sum_{i=1}^n d_B^2(\rho_i, \sigma_i) \right] \leq d_B^2(\rho^{(n)}, \sigma^{(n)}) \leq \sum_{i=1}^n d_B^2(\rho_i, \sigma_i). \quad (4.5)$$

$$1 - \exp \left[- \frac{1}{2} \sum_{i=1}^n d_{\text{tr}}^2(\rho_i, \sigma_i) \right] \leq d_{\text{tr}}(\rho^{(n)}, \sigma^{(n)}) \leq \sum_{i=1}^n d_{\text{tr}}(\rho_i, \sigma_i). \quad (4.6)$$

Proof. The fidelity is multiplicative under tensor products, $F(\rho^{(n)}, \sigma^{(n)}) = \prod_{i=1}^n F(\rho_i, \sigma_i)$. Also, by induction it is easily seen that $(1 - \prod_i x_i) \leq \sum_i (1 - x_i)$ for any collection of reals $x_i \in [0, 1]$. Thus:

$$d_B^2(\rho^{(n)}, \sigma^{(n)}) = 1 - \prod_{i=1}^n F(\rho_i, \sigma_i) \leq \sum_{i=1}^n (1 - F(\rho_i, \sigma_i)) = \sum_{i=1}^n d_B^2(\rho_i, \sigma_i).$$

Since $\prod_i e^{x_i-1} \geq \prod_i x_i$ whenever $x_i \geq 0$, we get the lower bound in (4.5):

$$d_B^2(\rho^{(n)}, \sigma^{(n)}) = 1 - \prod_{i=1}^n F(\rho_i, \sigma_i) \geq 1 - \prod_{i=1}^n e^{F(\rho_i, \sigma_i)-1} = 1 - \exp \left[- \sum_{i=1}^n d_B^2(\rho_i, \sigma_i) \right].$$

The upper bound in Eqn. (4.6) follows by a calculation similar to the one yielding Eqn. (4.2), while the lower bound follows from (4.5) and two of the inequalities in Eqn. (2.36). \square

4.3 Examples of the cutoff phenomenon and applications

Theorem 45 establishes a pre-cutoff and thereby estimates, up to a factor of 2, the time to convergence. The next natural question is: when does an actual cutoff occur? We discuss two such situations. The first concerns tensor powers of primitive channels where the input states are restricted to be separable, the second concerns tensor powers of channels whose unique fixed point is a pure state. In subsection 4.3.3 we give an explicit example of this situation (the qubit amplitude damping channel).

4.3.1 Primitive Channels with Separable Initial States

Beyond Theorem 45, we can establish a sharp cutoff for primitive Liouvillians when the inputs are restricted to be fully separable quantum states between the n channels:

$$\mathcal{S}_{\otimes_i d_i}^{\text{sep}} := \left\{ \sum_k p_k \rho_1^k \otimes \dots \otimes \rho_n^k \mid p_k \geq 0, \sum_k p_k = 1, \rho_i^k \in \mathcal{S}_{d_i} \right\}. \quad (4.7)$$

Proposition 47 (Primitive Liouvillians with separable inputs). *Let $\mathcal{L} : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be the generator, with gap $\bar{\lambda}$, of a one-parameter semigroup of primitive channels $T_t \equiv e^{t\mathcal{L}}$ ($t \geq 0$), and define the trace-norm convergence of $T_t^{(n)} \equiv T_t^{\otimes n}$ restricted to separable input states:*

$$\eta_{\text{tr}}^{\text{sep}}[T_t^{(n)}] := \sup_{\rho \in \mathcal{S}_{d^{\otimes n}}^{\text{sep}}} d_{\text{tr}}(T_t^{\otimes n}(\rho), T_{\varphi}^{\otimes n}(\rho)). \quad (4.8)$$

Then, the sequence of one-parameter semigroups $T_t^{(n)} \equiv T_t^{\otimes n}$ exhibits a cutoff (with respect to the convergence measure $\eta_{\text{tr}}^{\text{sep}}$) at times $t_n = \log(n)/2\bar{\lambda}$.

Proof. From the primitivity of the channel we get that $T_{\varphi}(\rho) = \sigma$ for any input state $\rho \in \mathcal{S}_d$, where σ is the unique stationary state of \mathcal{L} . Further, as σ is of full rank, the new bound from Proposition 14 and bounds from Eqn. (14) together with Theorem 20 show that, for any $\nu < \bar{\lambda}$, there exist constants $R > L > 0$ such that $Le^{-t\bar{\lambda}} \leq \eta_B[T_t] \leq Re^{-t\nu}$ for all $t \geq 0$.

The rest of the proof follows the same lines as the proof of Theorem 45. However, we first show the theorem here for the Bures metric, i.e. replacing d_{tr} in (4.8) by d_B . For proving the lower bound, the same arguments as the ones leading to Eqn. (4.3) show that $\lim_{n \rightarrow \infty} \eta_B^{\text{sep}}[T_{ct_n}] = 1$ for $c \in (0, 1)$. For the upper bound, note that due to convexity of the Bures distance (derived from concavity of the fidelity [NC00]) the supremum in (4.8) is reached for a product state $\rho = \bigotimes_{i=1}^n \rho_i$, so that Lemma 47 can be applied:

$$\begin{aligned} \left(\eta_B^{\text{sep}}[T_{ct_n}^{(n)}] \right)^2 &= \sup_{\rho_i \in \mathcal{S}_d} d_B^2 \left(\bigotimes_{i=1}^n T_{ct_n}(\rho_i), \sigma^{\otimes n} \right) \leq \sum_{i=1}^n \sup_{\rho_i \in \mathcal{S}_d} d_B^2(T_{ct_n}(\rho_i), \sigma) \\ &\leq nR^2 e^{-2c\nu t_n} = R^2 n^{1-c\nu/\bar{\lambda}}. \end{aligned}$$

As in the proof of Theorem 45, for each given $c > 1$ one can choose ν and R accordingly to show $\lim_{n \rightarrow \infty} \eta_B^{\text{sep}}[T_{ct_n}] = 0$, which proves a cutoff at times t_n . Finally, by Eqn. (2.36), a cutoff in the Bures convergence measure η_B^{sep} is equivalent to a cutoff in the trace-norm convergence $\eta_{\text{tr}}^{\text{sep}}$, at the same times t_n . \square

We do not know whether the separable input assumption is actually necessary for Proposition 47. If the assumption were indeed necessary, then the statement would imply the possibility of increased storage time of classical information due to an entangled encoding.

4.3.2 Channels with Unique Pure State Fixed Point

Proposition 48 (Unique pure state fixed point). *Suppose that the pure state $\psi = |\psi\rangle\langle\psi| \in \mathcal{S}_d$ is the unique stationary state of the Liouvillian \mathcal{L} , which has gap $\bar{\lambda}$ and generates the*

channels $T_t := e^{t\mathcal{L}}$ ($t \geq 0$). Then, $T_t^{(n)} \equiv T_t^{\otimes n}$ exhibits a trace-norm cutoff at times $t_n = \log(n)/\bar{\nu}$, for some $\bar{\lambda} \leq \bar{\nu} \leq 2\bar{\lambda}$.

Proof. Since \mathcal{L} has only one stationary state, the peripheral spectrum of $e^{t\mathcal{L}}$ is trivial for all $t > 0$, so that $T_\varphi(\rho) = \psi$ for all $\rho \in \mathcal{S}_d$. This, together with well-known inequalities relating the fidelity and the trace-norm between a pure and a mixed state [NC00], yields:

$$1 - \inf_{\rho \in \mathcal{S}_{d^n}} F^2(T_t^{\otimes n}(\rho), \psi^{\otimes n}) \leq \eta_{tr}[T_t^{\otimes n}] \leq \sqrt{1 - \inf_{\rho \in \mathcal{S}_{d^n}} F^2(T_t^{\otimes n}(\rho), \psi^{\otimes n})}. \quad (4.9)$$

The last infimum can be evaluated explicitly in the case at hand:

$$\begin{aligned} \inf_{\rho \in \mathcal{S}_{d^n}} F^2(T_t^{\otimes n}(\rho), \psi^{\otimes n}) &= \inf_{\rho \in \mathcal{S}_{d^n}} \text{tr}[T_t^{\otimes n}(\rho)\psi^{\otimes n}] = \inf_{\rho \in \mathcal{S}_{d^n}} \text{tr}[\rho (T_t^*(\psi))^{\otimes n}] \\ &= \lambda_{\min} \left((T_t^*(\psi))^{\otimes n} \right) = [\lambda_{\min}(T_t^*(\psi))]^n \\ &= [1 - \lambda_{\max}(\mathbb{1} - T_t^*(\psi))]^n = (1 - \|T_t^*(\mathbb{1} - \psi)\|_\infty)^n, \end{aligned}$$

where in the last step we used that T_t is trace-preserving ($T_t^*(\mathbb{1}) = \mathbb{1}$) and that T_t^* is a positive map. Thus, from equation (4.9) above:

$$1 - (1 - \|T_t^*(\mathbb{1} - \psi)\|_\infty)^n \leq \eta_{tr}[T_t^{\otimes n}] \leq \sqrt{1 - (1 - \|T_t^*(\mathbb{1} - \psi)\|_\infty)^n}. \quad (4.10)$$

Since $T_\varphi(\rho) = \psi$ for all $\rho \in \mathcal{S}_d$, we have $T_\varphi(A) = \psi \text{tr}[A]$ for all $A \in \mathcal{M}_d$, so that its dual is given by $T_\varphi^*(B) = \mathbb{1} \text{tr}[B\psi]$ for $B \in \mathcal{M}_d$. Thus:

$$\|T_t^*(\mathbb{1} - \psi)\|_\infty = \|(T_t^* - T_\varphi^*)(\psi)\|_\infty.$$

Now we consider this last equation: When one writes ψ as a linear combination of the generalized eigenvectors of \mathcal{L}^* , then considering large times t will essentially pick out the Jordan-eigenvalue(s) occurring in ψ which has largest real part $-\bar{\nu} < 0$ (i.e., not the eigenvalue 0), and among these it will pick out the polynomial(s) of highest degree $J \geq 0$ that are occupied (i.e., occur with non-zero coefficient in the linear decomposition of ψ). This means that, for any arbitrarily chosen $t_0 > 0$, there exist constants $0 < C_1 \leq C_2$ such that

$$C_1(\bar{\nu}t)^J e^{-\bar{\nu}t} \leq \|(T_t^* - T_\varphi^*)(\psi)\|_\infty \leq C_2(\bar{\nu}t)^J e^{-\bar{\nu}t} \quad \forall t > t_0.$$

Using this in Eqn. (4.10) gives

$$1 - \left(1 - C_1(\bar{\nu}t)^J e^{-\bar{\nu}t}\right)^n \leq \eta_{tr}[T_t^{\otimes n}] \leq \sqrt{1 - (1 - C_2(\bar{\nu}t)^J e^{-\bar{\nu}t})^n}.$$

This proves a cutoff for $T_t^{\otimes n}$ at times $t_n = (\log n)/\bar{\nu}$, since for any constants $c, K > 0$ and $J \geq 0$:

$$\begin{aligned} \lim_{n \rightarrow \infty} 1 - (1 - K(\bar{\nu}ct_n)^J e^{-\bar{\nu}ct_n})^n &= 1 - \lim_{n \rightarrow \infty} \left(1 + \frac{-K(c \log n)^J n^{1-c}}{n}\right)^n \\ &= 1 - \lim_{n \rightarrow \infty} \exp\left(-K(c \log n)^J n^{1-c}\right) = \begin{cases} 1, & c < 1 \\ 0, & c > 1. \end{cases} \end{aligned}$$

As $(-\bar{\nu})$ is the real part of an eigenvalue of \mathcal{L} , it is evident that $\bar{\nu} \geq \bar{\lambda}$, and Theorem 45 shows $\bar{\nu} \leq 2\bar{\lambda}$. \square

Both infima in the upper bound and in the lower bound in Eqn. (4.9) are attained for some pure *product* state $\rho = \varphi^{\otimes n}$, even though this is not clear for the supremum that achieves $\eta_{tr}[T_t^{\otimes n}]$. The paradigmatic example in the following subsection saturates the upper bound $\bar{\nu} = 2\bar{\lambda}$, but we also provide modifications of this example where $\bar{\nu}$ takes on any values between $\bar{\lambda}$ and $2\bar{\lambda}$.

4.3.3 Qubit Amplitude Damping

The amplitude damping process (on qubits) describes the situation where the excited state $|1\rangle$ decays into the ground state $|\psi\rangle := |0\rangle$ at a constant rate γ . This corresponds to a Master equation with a single Lindblad operator $L := \sqrt{\gamma}|0\rangle\langle 1|$ and no coherent contribution:

$$\mathcal{L}(\rho) := L\rho L^\dagger - \frac{1}{2}L^\dagger L\rho - \frac{1}{2}\rho L^\dagger L = \gamma \left(|0\rangle\langle 0| \cdot \langle 1|\rho|1\rangle - \frac{1}{2}|1\rangle\langle 1|\rho - \frac{1}{2}\rho|1\rangle\langle 1| \right). \quad (4.11)$$

A straightforward calculation shows that:

$$\eta_{tr}[e^{t\mathcal{L}}] = \begin{cases} e^{-\gamma t} & , 0 \leq t \leq (\log 2)/\gamma \\ e^{-\gamma t/2} / \sqrt{4(1 - e^{-\gamma t})} & , t \geq (\log 2)/\gamma. \end{cases}$$

The contraction maximum in the contraction measure is attained for $\varphi = |1\rangle$, when $0 \leq t \leq (\log 2)/\gamma$, and for $\varphi \propto |1\rangle + \sqrt{1 - 2e^{-\gamma t}}|0\rangle$ otherwise. Thus, $\eta_{tr}[e^{t\mathcal{L}}]$ decays asymptotically in time as $e^{-\gamma t/2}$, and *not* as $e^{-\gamma t}$, which one would expect from the analogous classical noise process (the classical Markov map has eigenvalues 0 and $-\gamma$, whereas the Liouvillian (4.11) has two additional eigenvalues $-\gamma/2$).

By Proposition 48, the semigroups $(e^{t\mathcal{L}})^{\otimes n}$ exhibit a trace-norm cutoff at times $t_n = (\log n)/\bar{\nu}$ for some $\bar{\nu}$ with $\bar{\lambda} = \gamma/2 \leq \bar{\nu} \leq 2\bar{\lambda} = \gamma$. $\bar{\nu}$ can be computed explicitly by using $\|T_t^*(\mathbb{1} - \psi)\|_\infty = e^{-\gamma t}$ in Eqn. (4.10), or from the proof of Proposition 48 by writing ψ as a linear combination of eigenvectors of \mathcal{L}^* :

$$\psi = |0\rangle\langle 0| = \mathbb{1} - |1\rangle\langle 1|, \quad (4.12)$$

where $\mathbb{1}$ and $|1\rangle\langle 1|$ are eigenvectors of \mathcal{L}^* with eigenvalues 0 and $-\gamma$, respectively. Thus, $\bar{\nu} = \gamma = 2\bar{\lambda}$, and the cutoff occurs at times $t_n = (\log n)/\gamma$.

For system size n it thus takes time $O(\log n)$ before convergence happens, even though the Liouvillians

$$\mathcal{L}^{(n)} = \mathcal{L} \otimes \text{id} \otimes \dots \otimes \text{id} + \text{id} \otimes \mathcal{L} \otimes \dots \otimes \text{id} + \dots + \text{id} \otimes \text{id} \otimes \dots \otimes \mathcal{L}, \quad (4.13)$$

which generate the semigroups $T_t^{(n)} = e^{t\mathcal{L}^{(n)}}$, have a gap $\bar{\lambda}^{(n)} = \bar{\lambda} = \gamma/2$ which is *independent* of n . Therefore, this example refutes the conventional wisdom whereby “the gap governs the convergence time”.

If, in addition to the Liouvillian (4.11), there are also processes with Lindblad operators $\sqrt{\alpha}|0\rangle\langle 0|$ and $\sqrt{\beta}|1\rangle\langle 1|$ acting on each qubit, then the steady state ψ and its decomposition (4.12) into eigenvectors of the dual evolution operator are as above (in particular,

$\bar{\nu} = \gamma$), and a cutoff occurs at times $t_n = (\log n)/\gamma$. In this new situation, however, the gap is given by $\bar{\lambda} = \min\{\gamma, (\gamma + \alpha + \beta)/2\}$, which shows that the bounds on the cutoff time given by Proposition 48 and implied by Theorem 45 are tight.

Dissipative Engineering

5.1 Dissipative state preparation

The reliable and efficient preparation of entangled states has been one of the main tasks in quantum information science since the birth of the field. The effort has been driven on the one hand by the desire to understand these quintessentially non-classical states of matter, and on the other, by their promise as building blocks for quantum information processing tasks. In particular, bipartite maximally entangled states constitute the gold standard of entanglement theory, which in turn is believed to be the main ingredient responsible for the additional information processing power of quantum machines over classical ones. As maximally entangled states are such an important resource in many quantum information processing protocols (ex: repeaters, cryptography), having access to a reliable source of them cannot be overestimated. Since the advent of quantum information science, noise has been considered a detrimental element in a physical setup, causing decoherence which must at all cost be avoided. A few years ago, however, it has been suggested that dissipative noise can be used as a resource for quantum information processing, abetting in the preparation of entangled states [VWC09, DMK⁺08, KBD⁺08].

The first experimental studies along these lines [LHN⁺11, KMJ⁺11] have shown these new ideas to be realistic and promising as a new path for harnessing the potential of quantum information. In this section, we consider two Λ -atoms trapped in a single mode cavity QED setup [PCZ96, DRBH95] coherently driven by a classical optical field and a microwave or Raman field. We demonstrate that a maximally entangled stationary state of the two atoms can be prepared dissipatively with very high fidelity. In this scheme, the two atoms are rapidly driven into a singlet state, independent of the initial state of the system, and without need for any unitary feedback control. Consequently, the lifetime of the state is dictated by the lifetime of the experiment. We identify the relevant interactions by systematically truncating the Hilbert space of the problem using an effective operator formalism based on second order perturbation theory of the excited states. This gives us an effective master equation from which all of the desired performance measures can be analytically derived. In particular, we analyze the optimal stationary-state fidelity and the convergence time as a function of system parameters. We show that the fidelity of our scheme scales quadratically better in the cooperativity (the invariant quality measure of the cavity QED setup) than any known coherent unitary protocol. Our analysis thus

indicates that dissipative state preparation is more than just a new conceptual approach, but can allow for significant improvement as compared to preparation protocols based on coherent unitary dynamics.

Attached:

Dissipative preparation of entanglement in optical cavities

M. J. Kastoryano, F. Reiter, A. S. Sørensen,

Phys. Rev. Lett. **106**, 090502 (2011).

Dissipative Preparation of Entanglement in Optical Cavities

M. J. Kastoryano,^{1,2} F. Reiter,¹ and A. S. Sørensen¹

¹QUANTOP, Danish Quantum Optics Center, Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

²Niels Bohr International Academy, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

(Received 5 November 2010; published 28 February 2011)

We propose a novel scheme for the preparation of a maximally entangled state of two atoms in an optical cavity. Starting from an arbitrary initial state, a singlet state is prepared as the unique fixed point of a dissipative quantum dynamical process. In our scheme, cavity decay is no longer undesirable, but plays an integral part in the dynamics. As a result, we get a qualitative improvement in the scaling of the fidelity with the cavity parameters. Our analysis indicates that dissipative state preparation is more than just a new conceptual approach, but can allow for significant improvement as compared to preparation protocols based on coherent unitary dynamics.

DOI: 10.1103/PhysRevLett.106.090502

PACS numbers: 03.67.Bg, 42.50.Dv, 42.50.Pq

Preparing entangled states faithfully and reliably has been one of the major challenges in the field of experimental quantum information science, where a plethora of different systems has been investigated [1]. In particular, several schemes based on cavity QED have been proposed (see, e.g., [2–8]), and these schemes have been used to generate entanglement of atoms using microwave cavities [9,10].

Traditionally, it has been assumed that noise can only have detrimental effects in quantum information processing. Recently, however, it has been suggested [11–14], and realized experimentally [15], that the environment can be used as a resource. In particular, it was shown in Ref. [11] that universal quantum computation is possible using only dissipation, and that a very large class of states, known as tensor product states [16,17], can be prepared efficiently. On general grounds, one may argue that dissipative state preparation can have significant advantages over other state preparation methods by converting a detrimental source of noise into a resource. Whether this is really true can, however, only be determined by considering concrete physical systems. To answer this question we study the generation of entanglement in high finesse optical cavities [18–20]. Generating entanglement in this system by unitary evolution has been studied in great detail theoretically (see, e.g., [2–5]) and the limitations coming from dissipation are thus well understood. We find that dissipative state preparation leads to higher fidelity entangled states than schemes based on unitary dynamics for this system. Our results thus indicate that dissipative state preparation is more than just a new conceptual approach, but may also be of significant advantage in practice.

In this Letter, we suggest a dissipative scheme for preparing an entangled state of two Λ atoms in an optical cavity, with detunings as depicted in Fig. 1(a). Our scheme can be understood from Fig. 1(b), which describes the effective coupled ground states of the atoms in the cavity. A microwave field shuffles the three triplet state around while the cavity interaction causes a transition $|00\rangle|0\rangle \rightarrow |S\rangle|1\rangle$ followed by a rapid decay to $|S\rangle|0\rangle$. The latter is

coupled to $|11\rangle|1\rangle$ which then decays to $|11\rangle|0\rangle$. Here the first ket in the pair refers to the atoms, the second to the cavity photon number, and $|S\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$ is the singlet state. The first cavity transition $|00\rangle|0\rangle \rightarrow |S\rangle|1\rangle$ is shifted by g^2/Δ due to the interaction of the photon with a single atom in state $|1\rangle$ in the final state $|S\rangle|1\rangle$, while the second transition $|S\rangle|0\rangle \rightarrow |11\rangle|1\rangle$ is shifted by twice that amount, $2g^2/\Delta$, due to the interaction of the photon with two atoms in state $|1\rangle$. Setting the cavity detuning equal to g^2/Δ will greatly favor the transition to the singlet state and strongly suppress the transition away from it. Thus, essentially all of the population is driven into the maximally entangled singlet state.

Our protocol actively exploits the cavity decay to drive the system to a maximally entangled stationary state. The only generic source of noise left in the system is then the one coming from spontaneous emission. This leads, quite remarkably, to a linear scaling of the fidelity with the cooperativity [see the inset in Fig. 1(c)], which is in contrast to schemes based on controlled unitary dynamics, where there are two malevolent noise sources, cavity and atomic decay, typically resulting in a weaker square root scaling of the fidelity [2–6].

We point out that a similar study to ours has been conducted by Wang and Schirmer [21], where they consider a detuning of the energy levels in order to break the symmetry in the system, and guarantee a unique steady state. It can be shown that their scheme, when adapted to optical cavities, does not give a linear scaling of the fidelity [22], but rather the square root, as for coherent unitary protocols.

In the following, the system-environment interaction will be assumed Markovian, and can thus be modeled by a master equation in Lindblad form:

$$\dot{\rho} = i[\rho, H] + \sum_j L_j \rho L_j^\dagger - \frac{1}{2}(L_j^\dagger L_j \rho + \rho L_j^\dagger L_j), \quad (1)$$

where the L_j 's are the so-called Lindblad operators. We derive a master equation for the ground states of the

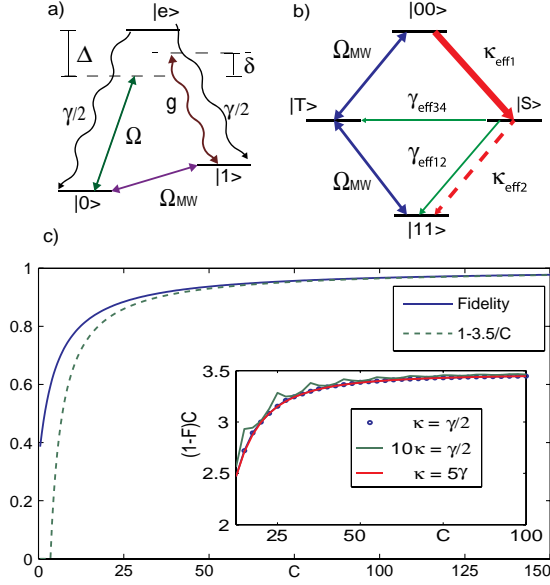


FIG. 1 (color online). (a) Level diagram of a single atom with laser detuning Δ and cavity detuning δ from two photon resonance. The optical pumping laser for the two atoms differs by a relative phase of π . (b) The effective two qubit system. The driving Ω_{MW} causes rapid transitions between the three triplet states. The atoms decay through the cavity from $|00\rangle$ to $|S\rangle$ and from $|S\rangle$ to $|11\rangle$ with effective decay rates $\kappa_{eff,1}$ and $\kappa_{eff,2}$, where $\kappa_{eff,1} \gg \kappa_{eff,2}$. The spontaneous emission rates $\gamma_{eff,i}$ will tend to reduce the fidelity by redistributing information to the triplet states. (c) Fidelity as a function of the cooperativity $C = g^2/\kappa\gamma$. The inset gives a more accurate account of the scaling ($1 - F \leq 3.5C^{-1}$) for different values of κ/γ .

system, which has the singlet as unique stationary state. This is achieved in our setup, by constructing an effective master equation, whose main contributing terms are the Hamiltonian $H = \frac{1}{2}\Omega_{MW}(J_+ + J_-)$, and the Lindblad operator $L^\kappa = \sqrt{\kappa_{eff}}|S\rangle\langle 00|$, where $J_+ = |1\rangle\langle 0| \otimes \mathbb{1} + \mathbb{1} \otimes |1\rangle\langle 0|$. It can readily be seen that the microwave field (Ω_{MW}) drives the transitions between the three triplet states ($\{|00\rangle, |11\rangle, |T\rangle = (|01\rangle + |10\rangle)/\sqrt{2}\}$), while the Lindblad operator, originating from the cavity field leakage, will drive the transitions from $|00\rangle$ to $|S\rangle$. The singlet state is thus the unique fixed point of this system, and has a relaxation rate $\sim \min\{\Omega_{MW}^2/\kappa_{eff}, \kappa_{eff}\}$. Other terms will contribute weakly to the dynamics of the system, and slightly perturb the stationary state away from $|S\rangle$.

We now show how to prepare this effective system in a realistic quantum optical setup. Our setup, shown in Fig. 1(a), consists of two Λ -type three level atoms in a detuned cavity with two stable lower energy states $|0\rangle$ and $|1\rangle$, and an excited state $|e\rangle$ with a large energy separation to the lower lying states. We apply one far off-resonance optical laser, with detuning Δ , driving the $0 \leftrightarrow e$ transition and a microwave field driving the $0 \leftrightarrow 1$ transition resonantly. The cavity mode couples the $1 \leftrightarrow e$ transition off-resonantly, with detuning $\Delta - \delta$, where δ is the cavity detuning from two photon resonance. Furthermore, we

assume a π phase difference in the optical laser between the two atoms. This phase difference is crucial in guaranteeing that the singlet is the unique stationary state of the reduced system.

In a rotating frame, this situation is described by the Hamiltonian $H = H_0 + H_g + V_+ + V_-$,

$$H_0 = \delta a^\dagger a + \Delta(|e\rangle_1\langle e| + |e\rangle_2\langle e|) + [g(|e\rangle_1\langle 1| + |e\rangle_2\langle 1|)a + \text{H.c.}], \quad (2)$$

$$H_g = \frac{\Omega_{MW}}{2}(|1\rangle_1\langle 0| + |1\rangle_2\langle 0|) + \text{H.c.}, \quad (3)$$

$$V_+ = \frac{\Omega}{2}(|e\rangle_1\langle 0| - |e\rangle_2\langle 0|), \quad (4)$$

where $V_- = V_+^\dagger$, g is the cavity coupling constant, a is the cavity field operator, Ω represents the optical laser driving strength, and Ω_{MW} the microwave driving strength. On top of the Hamiltonian dynamics, two sources of noise will inherently be present: spontaneous emission of the excited state of the atoms to the lower states with decay rates γ_i ; and cavity leakage at a rate κ . We assume for convenience that the spontaneous emission rates are the same for decaying to the $|0\rangle$ and to the $|1\rangle$ states (i.e., $\gamma_0 = \gamma_1 = \gamma/2$). This translates into five Lindblad operators governing dissipation $L^\kappa = \sqrt{\kappa}a$, $L_1^\gamma = \sqrt{\gamma/2}|0\rangle_1\langle e|$, $L_2^\gamma = \sqrt{\gamma/2}|0\rangle_2\langle e|$, $L_3^\gamma = \sqrt{\gamma/2}|1\rangle_1\langle e|$, $L_4^\gamma = \sqrt{\gamma/2}|1\rangle_2\langle e|$.

If the optical pumping laser is sufficiently weak, and if the excited states are not initially populated, then the excited states of the atoms, as well as the excited cavity field modes, can be adiabatically eliminated. The resulting effective dynamics will describe two two-level systems in a strongly dissipative environment. To second order in perturbation theory, the dynamics are then given by the effective operators [22]:

$$H_{eff} = -\frac{i}{2}[V_- H_{NH}^{-1} V_+ + V_+ (H_{NH}^{-1})^\dagger V_-] + H_g, \quad (5)$$

$$L_{eff,j} = L_j H_{NH}^{-1} V_+, \quad (6)$$

where $H_{NH} = H_0 - \frac{i}{2}\sum_j L_j^\dagger L_j$ is a non-Hermitian Hamiltonian describing the nonunitary dynamics of the excited states which we eliminate. Applying the above equations to our setup, and keeping only terms to lowest order in Ω , the operators in the effective Master equation can be evaluated explicitly, yielding the effective Hamiltonian and principle Lindblad operator

$$H_{eff} = \frac{1}{2}\Omega_{MW}(|1\rangle_1\langle 0| + |1\rangle_2\langle 0| + \text{H.c.}) + \mathcal{O}\left(\frac{\Omega^2}{\Delta}\right) \quad (7)$$

$$L_{eff}^\kappa = \sqrt{\frac{g_{eff}^2 \kappa/2}{(g^2/\Delta - \delta)^2 + (\kappa/2 + \gamma\delta/2\Delta)^2}}|S\rangle\langle 00| + \sqrt{\frac{g_{eff}^2 \kappa/2}{(2g^2/\Delta - \delta)^2 + (\kappa/2 + \gamma\delta/2\Delta)^2}}|11\rangle\langle S|, \quad (8)$$

where $g_{eff} = g\Omega/\Delta$.

The principal Lindblad operator L_{eff}^{κ} describes the decay from $|00\rangle$ to $|S\rangle$ at a rate $\kappa_{\text{eff},1}$ and from $|S\rangle$ to $|11\rangle$ at a rate $\kappa_{\text{eff},2}$. The effective decay rates $\kappa_{\text{eff},1}$ ($\kappa_{\text{eff},2}$), equal to the square of the first coefficient, are much smaller than the cavity decay $\kappa_{\text{eff},i} \ll \kappa$, such that the two decays happen sequentially. The first term in the denominators represents the effective detuning of the cavity which is shifted by g^2/Δ by each atom in state $|1\rangle$ in the final state. Setting the cavity detuning equal to the cavity line shift from a single atom and ensuring that this is much larger than the cavity loss ($g^2/\Delta = \delta \gg \kappa + \gamma\delta/\Delta$) strongly suppresses the transition out of the singlet state $\kappa_{\text{eff},2} \ll \kappa_{\text{eff},1} =: \kappa_{\text{eff}}$, as can be read off directly from Eq. (9). The effective Lindblad operator originating from cavity leakage then becomes $L_{\text{eff}}^{\kappa} \approx \sqrt{\kappa_{\text{eff}}} |S\rangle\langle 00|$. The effective Hamiltonian shuffles the triplet states among each other, so that the combined effect of the unitary and dissipative dynamics drives essentially all of the population to the singlet state. Hence, we have constructed an effective master equation which approximates the ideal situation described earlier.

We now consider imperfections imposed by spontaneous emission. The four Lindblad operators describing spontaneous emission will also transform into four independent effective noise operators for the reduced system. In the regime discussed above, and keeping only terms which drive the population out of the singlet state, the effective operators for spontaneous emission are

$$\begin{aligned} L_{\text{eff},i=1,2}^{\gamma} &= \sqrt{\gamma_{\text{eff},i=1,2}} |11\rangle\langle S| \\ L_{\text{eff},i=3,4}^{\gamma} &= \sqrt{\gamma_{\text{eff},i=3,4}} |T\rangle\langle S|, \end{aligned} \quad (9)$$

where $\gamma_{\text{eff}} = \gamma\Omega^2/2\Delta^2$, $\gamma_{\text{eff},i=1,2} = \gamma_{\text{eff}}/8$, and $\gamma_{\text{eff},i=3,4} = \gamma_{\text{eff}}/16$.

In order to evaluate the performance of the protocol we denote by P_j the probability to be in state j , and consider the rate of entering and exiting the singlet state $\dot{P}_S = P_{00}\kappa_{\text{eff},1} - P_S(\kappa_{\text{eff},2} + \sum_i \gamma_{\text{eff},i})$. The effective spontaneous emission will tend to modify the steady state population of the three triplet states, but we assume that Ω_{MW} is large compared to γ_{eff} and κ_{eff} , so that all three triplet states are almost equally populated in steady state. Solving for the stationary state of the rate equation and plugging in the decay rates obtained from the effective operators, we get for $P_S \approx 1$

$$\begin{aligned} 1 - F &\approx 3P_{00} \approx 12(3\gamma/16 + \kappa\Delta^2/2g^2) \\ &\times (\kappa/2 + \gamma g^2/2\Delta^2)/g^2\kappa. \end{aligned} \quad (10)$$

Here, F is the fidelity, which is given by the overlap of the stationary state of the dynamical process with the singlet state $F = |\langle S|\rho_{ss}|S\rangle| = P_S$.

There is a trade-off between the second term in the numerator ($\Delta^2\kappa/2g^2$) reflecting the probability to generate a cavity photon by decaying out of the singlet state, and the term in the denominator ($\gamma g^2/2\Delta$) reflecting the

scattering of cavity photons off the atoms. The first terms favors a small detuning Δ to increase the cavity line shift, whereas the second term favors a large detuning Δ to decrease the scattering. The optimal fidelity is reached when the two terms in the numerator are similar ($\gamma \approx \kappa\Delta^2/2g^2$), in which case the two terms in the sum in the denominator are also similar ($\kappa \approx \gamma 2g^2/\Delta^2$). This leads to an error scaling as

$$1 - F \propto C^{-1}, \quad (11)$$

where $C = g^2/\kappa\gamma$ is commonly referred to as the cooperativity. Plugging in the values in Eq. (10) one gets a proportionality factor of roughly 3 in Eq. (11). By numerically extracting the fixed point of the full master equation, and then maximizing its fidelity with respect to the singlet for fixed values of C , we get that the actual constant is closer to 3.5 [inset in Fig. 1(c)], i.e., $1 - F \approx 3.5C^{-1}$. This discrepancy can be attributed to the fact that we did not include all of the spontaneous emission terms in Eq. (10). In addition, the assumption that all three triplet states are equally populated is not exact. To support our analysis further, we also note that the fidelity scaling is essentially independent of the ratio κ/γ [inset in Fig. 1(c)].

For comparison, in a controlled unitary dynamics protocol, the fidelity will suffer errors coming from spontaneous emission on the one hand, and from cavity decay on the other. Decreasing one of the error sources will typically increase the other in such a way that the optimal value of the fidelity is $1 - F \propto 1/\sqrt{C}$ [6]. Indeed, to the best of our knowledge, all entangled state preparation protocols based solely on controlled unitary dynamics scale at best as $1/\sqrt{C}$ [2–5]. This means that the linear scaling of the fidelity from Eq. (11), is a quadratic improvement as compared to any known closed system entanglement preparation protocol. We note, however, that it is possible to beat this if one exploits measurement and feedback [6–8]. As mentioned previously, the reason for this improvement stems from the fact that cavity decay is used as a resource in our dissipative scheme, so that the only purely detrimental source of noise is the spontaneous emission. We point out as well that some systems, such as Circuit QED [10,23], are ill suited for measurement feedback schemes, as single photon detection can be a severe experimental hurdle. For such systems, it could very well be that a dissipative scheme is more favorable in practice.

The above analysis has been conducted without any consideration of the speed of convergence. We now show, that the entangled stationary state can be reached rapidly. In Fig. 2, we simulate the dynamics of the full master equation for an appropriate set of parameters. Starting from an arbitrary initial state, the populations of the triplet states undergo rapid coherent oscillations with an envelope decaying at a rate proportional to the gap (the smallest nonvanishing real part of an eigenvalue of the Liouvillian), while the singlet state converges to its maximum value at the same rate.

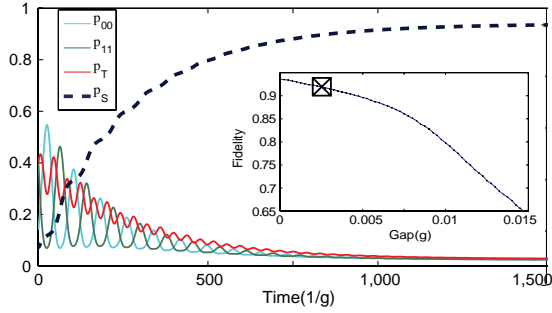


FIG. 2 (color online). The main figure shows the population of the singlet state (thick dashed line) and of the three triplet states (full lines) as a function of time for a random initial state. The curves were plotted for $C = 50$, $\kappa = \gamma/2$, $\Omega = 5 \Omega_{\text{MW}}/2$, $g = 20 \Omega$, and Δ, δ are such that they maximize the fidelity for small Ω . In this parameter regime, the stationary state has a 92% fidelity with respect to the singlet state. The inset shows the maximal fidelity as a function of the gap size for $C = 50$ and $\kappa = \gamma/2$. The main figure corresponds to the cross on the curve in the inset.

For a given cavity, g , κ , and γ can be considered fixed by experimental constraints, and the speed of convergence is primarily governed by the magnitudes of Ω and Ω_{MW} . The speed of convergence can be increased by increasing the driving laser strength (Ω), but the latter can not be too large otherwise perturbation theory breaks down, and the excited cavity and atomic states can no longer be ignored. Furthermore, Ω_{MW} can not be too small with respect to $\{\kappa_{\text{eff}}, \gamma_{\text{eff}}\}$, otherwise the coherent shuffling of the triplet states will not be sufficiently strong to keep them at equal population. The inset in Fig. 2 shows how the maximal fidelity scales as a function of the gap for a specific set of cavity parameters. The curve is plotted by optimizing the fidelity, for given fixed values of the gap, with respect to $\{\Omega, \Omega_{\text{MW}}\}$, for fixed values of $\{\Delta, \delta\}$ (those which are optimal for small Ω). There is clearly a trade-off between the accuracy of the dissipative state preparation protocol and the speed at which one reaches the stationary state, but close to optimal fidelity the dependence is weak.

Present day experimentally achievable values for the cooperativity are around $C \approx 30$ [18–20]. This puts our scheme at $\sim 90\%$ fidelity with respect to the singlet state. While this is still limited, the prospect for improving it is much more promising with the current protocol than for protocols based on controlled unitary dynamics; e.g., decreasing the error by an order of magnitude would require improving the cavity finesse by a factor of 10 as opposed to a factor of 100 with the square root scaling. Figure 2 shows that the stationary state is reached in a time $\sim 1000/g$, which yields for $g = (2\pi)35$ MHz [18] a convergence time of roughly $5 \mu\text{s}$ starting from an arbitrary initial state. This is much faster than typical decoherence time scales for this system.

We have investigated the possible advantage of dissipative state preparation by proposing a novel scheme for the

preparation of an entangled state of two trapped atoms in an optical cavity. From both analytical and numerical evidence, we give the scaling of the error explicitly, and show that the stationary state is reached rapidly. Our results indicate that not only can one produce entanglement dissipatively in a simple cavity system, but, to the best of our knowledge, the scaling of the fidelity for such entanglement preparation is better than any existing coherent unitary protocol. These results are an indication that an approach based on dissipation can be very fruitful for state preparation, as one manifestly can transform a previously undesirable noise source into a resource. It would be interesting to see if one could obtain similar results in related systems such as trapped ions or solid state based quantum devices, where dissipation traditionally plays a detrimental role.

We thank M. M. Wolf and D. Witthaut for helpful discussions. We acknowledge financial support from the European project QUEVADIS and from the Villum Kann Rasmussen Foundation.

- [1] T. D. Ladd *et al.*, *Nature (London)* **464**, 45 (2010).
- [2] T. Pellizzari, S. A. Gardiner, J. I. Cirac, and P. Zoller, *Phys. Rev. Lett.* **75**, 3788 (1995).
- [3] J. Pachos and H. Walther, *Phys. Rev. Lett.* **89**, 187903 (2002).
- [4] P. Domokos, J. M. Raimond, M. Brune, and S. Haroche, *Phys. Rev. A* **52**, 3554 (1995).
- [5] S.-B. Zheng and G.-C. Guo, *Phys. Rev. Lett.* **85**, 2392 (2000).
- [6] A. S. Sørensen and K. Mølmer, *Phys. Rev. Lett.* **91**, 097905 (2003).
- [7] M. B. Plenio, S. F. Huelga, A. Beige, and P. L. Knight, *Phys. Rev. A* **59**, 2468 (1999).
- [8] A. Beige, D. Braun, B. Tregenna, and P. L. Knight, *Phys. Rev. Lett.* **85**, 1762 (2000).
- [9] A. Rauschenbeutel *et al.*, *Science* **288**, 2024 (2000).
- [10] L. DiCarlo *et al.*, *Nature (London)* **467**, 574 (2010).
- [11] F. Verstraete, M. M. Wolf, and J. I. Cirac, *Nature Phys.* **5**, 633 (2009).
- [12] B. Kraus *et al.*, *Phys. Rev. A* **78**, 042307 (2008).
- [13] S. Diehl *et al.*, *Nature Phys.* **4**, 878 (2008).
- [14] G. Vacanti and A. Beige, *New J. Phys.* **11**, 083008 (2009).
- [15] H. Krauter *et al.*, *arXiv:1006.4344*.
- [16] M. Fannes, B. Nachtergaele, and R. F. Werner, *Commun. Math. Phys.* **144**, 443 (1992).
- [17] F. Verstraete and J. I. Cirac, *arXiv:cond-mat/0407066v1*.
- [18] R. Miller *et al.*, *J. Phys. B* **38**, S551 (2005).
- [19] A. D. Boozer, A. Boca, R. Miller, T. E. Northup, and H. J. Kimble, *Phys. Rev. Lett.* **97**, 083602 (2006).
- [20] A. Kubanek *et al.*, *Phys. Rev. Lett.* **101**, 203602 (2008).
- [21] X. Wang and S. G. Schirmer, *arXiv:1005.2114v2*.
- [22] F. Reiter, M. J. Kastoryano, and A. S. Sørensen (to be published).
- [23] D. Schuster *et al.*, *Nature (London)* **445**, 515 (2007).

5.1.1 Dissipative Preparation of Graph States

In this section, we consider the dissipative preparation of graph states. This task was considered in [VWC09, KBD⁺08], where it was shown that a set of local Lindblad operators can be constructed in a way analogous to Eqn. (4.13), such that the unique stationary state of the process is the desired graph state and that the spectral gap of the process is independent of the number n of particles or stabilizer operators.

We complete this analysis by showing that the convergence time, measured in trace-norm, scales as $\log n$ with the system size. Trace-norm convergence is the relevant quantity to consider in this case, as it quantifies the maximal failure probability when the graph state is used for further quantum information processing, like the cluster state for measurement-based quantum computing. We actually show that the dissipative preparation of graph states exhibits a cutoff (in trace-norm) at times of order $O(\log n)$ in the sense of Definition 42. Although still efficient, the $\log n$ scaling of the preparation time again indicates that the gap does not fully determine the convergence behavior.

Proposition 49 (Dissipative preparation of graph states). *Any graph state on n sites, associated to a graph of maximal degree k , can be prepared dissipatively in a time of order $\log n$ by using n Lindblad operators that are at most k -local.*

In fact, for large n , the preparation procedure described in [VWC09, KBD⁺08] takes exactly time $t_n = (\log n) / \gamma$ to converge to the desired graph state, where γ is the decay rate ($\gamma/2 = \text{spectral gap}$) of each local Lindblad operator and when starting from the most disadvantageously chosen initial state.

Proof. Given a set $\{S_k\}_{k=1}^n$ of stabilizer operators, the unique state which is an eigenstate of S_k with eigenvalue $+1$ for every k is called a stabilizer state. Graph states [HDE⁺05] are a special case of these and can be described by an undirected graph with n vertices. The stabilizer operators of the graph state are then $S_k = \sigma_k^x \prod_{j \in \text{nbhd}(k)} \sigma_j^z$, where $\text{nbhd}(k)$ denotes the set of all vertices connected to vertex k by an edge.

The stabilizer operators of a graph state uniquely defines a “graph basis”, written as $\{|\Phi_{i_1, \dots, i_n}\rangle\}_{i_l \in \{0,1\}}$, by $S_k |\Phi_{i_1, \dots, i_n}\rangle = (-1)^{i_k} |\Phi_{i_1, \dots, i_n}\rangle$. These basis vectors satisfy $\sigma_k^z |\Phi_{i_1, \dots, i_k=1, \dots, i_n}\rangle = |\Phi_{i_1, \dots, i_k=0, \dots, i_n}\rangle$, and the “graph state” is $|\Phi_{0, \dots, 0}\rangle$.

Define the n Lindblad operators [KBD⁺08] ($k = 1, \dots, n$)

$$L_k = \sqrt{\gamma} \sigma_k^z \frac{\mathbb{1} - S_k}{2}, \quad (5.1)$$

and observe that $L_k |\Phi_{i_1, \dots, i_k=1, \dots, i_n}\rangle = \sqrt{\gamma} |\Phi_{i_1, \dots, i_k=0, \dots, i_n}\rangle$ and $L_k |\Phi_{i_1, \dots, i_k=0, \dots, i_n}\rangle = 0$. Thus, in the graph basis, each of these Lindblad operators acts as one term of the sum (4.13) acts in the computational basis. Therefore, together they act like the tensor product of amplitude damping channels in subsection 4.3.1, now with the graph state as the stationary state. Proposition 48 or, more explicitly, subsection 4.3.1 thus prove a cutoff at times $(\log n) / \gamma$ for the preparation of graph states. \square

Note in particular, Proposition 49 shows that, for the procedure described by Eqn. (5.1), there exist some initial states for which one can guarantee convergence not to occur before time $(\log n) / \gamma$.

5.2 Dissipative computation

This section can be seen as a detailed exposition of the proof in [VWC09] that dissipative quantum computation can efficiently simulate circuit quantum computation.

5.2.1 Dissipative quantum computation on $\mathcal{H}_2^{\otimes N} \otimes \mathcal{H}_{M+1}$ with quasi-local Lindblad operators

Consider an N qubit quantum circuit consisting of a sequence of $M \in \mathbb{N}$ 2-local unitary gates $\{U_t\}_{t=1}^M$. Suppose that the input of the computation is encoded in the circuit, and define an initial reference state of the circuit $|0\rangle^{\otimes N}$, so that the intermediate state of the computation after a “time” $t \leq M$ is

$$|\psi_t\rangle = U_t U_{t-1} \dots U_1 |0\rangle^{\otimes N}, \quad (5.2)$$

where $|\psi_M\rangle$ is the final, desired, state of the computation. It is understood that an efficient computation is one for which M scales as $\text{poly}(N)$.

Our goal is to prepare an open quantum system, accurately modeled by a time-independent Markovian master equation, which relaxes rapidly¹ to a unique stationary state from which $|\psi_M\rangle$ can be read off efficiently. In other words, we will prepare a time-independent master equation $\dot{\rho} = \mathcal{L}(\rho)$ with a Liouvillian in Lindblad form

$$\mathcal{L}(\rho) = i[\rho, H] + \sum_k L_k \rho L_k^\dagger - \frac{1}{2} \{L_k^\dagger L_k, \rho\}. \quad (5.3)$$

We want to construct local $\{L_k, H\}$ which guarantee that (i) the master equation has a unique stationary state ρ_{ss} , (ii) ψ_M can be read off from ρ_{ss} in a time $\text{poly}(N, M)$, (iii) the relaxation time of the semigroup is of order $\text{poly}(N, M)$.

As in Feynman’s construction of a quantum simulator, we consider a Hilbert space split into a logical part consisting of N qubits, and a *time register* with states $\{|t\rangle\}_{t=0}^M$. We define the Lindblad operators of the system:

$$L_i = \sqrt{\gamma} |0\rangle_i \langle 1| \otimes |0\rangle_t \langle 0| \quad (5.4)$$

$$L_\alpha = \sqrt{\gamma} (U_\alpha \otimes |\alpha+1\rangle \langle \alpha| + U_\alpha^\dagger \otimes |\alpha\rangle \langle \alpha+1|), \quad (5.5)$$

where $i = 1, \dots, N$ and $\alpha = 0, \dots, M^2$. $|0\rangle_i \langle 1|$ is short-hand notation for $\text{id}_{2^{i-1}} \otimes |0\rangle \langle 1| \otimes \text{id}_{2^{N-i}}$, whereas the t subscript in (5.4) is just meant as a reminder that we are referring to the time registry. Our construction can be described purely dissipatively; i.e. we let $H = 0$. We assume for simplicity that the frequency (γ) of the two Lindblad operators is the same, but this is not necessary. Clearly, the Lindblad operators act locally on the qubits, but non-locally on the time register. Later, we will show that the time register can be efficiently encoded locally as well.

¹Provided the computation was efficient.

²Throughout this section, the logical subspaces will be indexed by a latin letter (usually i), while the time degrees of freedom will be indexed by a greek letter (usually α).

It is easy to check that

$$\rho_{ss} = \frac{1}{M+1} \sum_{t=0}^M |\psi_t\rangle\langle\psi_t| \otimes |t\rangle\langle t| \quad (5.6)$$

is a stationary state of the master equation. It is also the unique stationary state of the system, because Eqns. (5.4) and (5.5) satisfy the Davies-Frigiero-Spohn criterion (11). Later, by analyzing the spectrum of $\hat{\mathcal{L}}$, we will also see directly that the fixed point is unique. The outcome of the computation can manifestly be extracted from ρ_{ss} with probability $1/(M+1)$ by measuring the time register. Therefore, all that is left to show is that the semigroup relaxes in a time which is $\text{poly}(N, M)$.

We now state the main theorem of this section:

Theorem 50. *The quantum Markov semigroup defined by the Lindblad operators of Eqns. (5.4), (5.5) converges to its unique stationary state in a time which scales, at worst, as $t_{mix} = \mathcal{O}(NM^3 \log M)$.*

For now, we prove the theorem for the case of an $(M+1)$ -state time register. In the next section, we will show that the time register can be replaced by the unary encoding of Kitaev, without changing the order of the convergence time.

It turns out to be more convenient to work in the Linear operator representation. By Proposition 18, we have

$$2\eta[e^{t\hat{\mathcal{L}}}] = \sup_{\phi \in \mathcal{S}_d^+} \|e^{t\hat{\mathcal{L}}}(\phi) - \rho_s\|_1 \leq \sqrt{2d} \|e^{t\hat{\mathcal{L}}} - \hat{T}_\infty\| \quad (5.7)$$

The dimensional factor of $\sqrt{d} = 2^{N/2} \sqrt{M+1}$ translates to a factor of $\mathcal{O}(N \log M)$ in the convergence time bound.

In the linear operator picture, and using the canonical Matrix basis, the Liouvillian is given by $\hat{\mathcal{L}} = \hat{\mathcal{L}}_{\text{init}} + \hat{\mathcal{L}}_{\text{comp}}$, where, given that the Lindblad operators are real,

$$\hat{\mathcal{L}}_{\text{init}} = \sum_{i=1}^N L_i \otimes L_i - \frac{1}{2} (L_i^\dagger L_i \otimes \mathbb{1} + \mathbb{1} \otimes L_i^\dagger L_i) \quad (5.8)$$

$$\hat{\mathcal{L}}_{\text{comp}} = \sum_{\alpha=0}^M L_\alpha \otimes L_\alpha - \frac{1}{2} (L_\alpha^\dagger L_\alpha \otimes \mathbb{1} + \mathbb{1} \otimes L_\alpha^\dagger L_\alpha) \quad (5.9)$$

In order to bound the convergence time, we first estimate the gap of $\hat{\mathcal{L}}$, and then show that the prefactor cannot be too large. The trick, in bounding the gap, is to apply two successive similarity transformations to this Liouvillian, which bring it into a more tractable form. The first serves to eliminate the unitaries in Eqn. (5.4), and the second serves to bring the logical subsystem into diagonal form.

Consider the unitary operator

$$W = \sum_{\alpha} U_{\alpha-1} \dots U_1 \otimes |\alpha\rangle_t \langle \alpha|, \quad (5.10)$$

then the unitary transformation $W^\dagger \otimes W^\perp \mathcal{L} W \otimes \bar{W}$ leaves $\hat{\mathcal{L}}_{\text{init}}$ unchanged, but reduces the Lindblad operators in $\mathcal{L}_{\text{comp}}$ to $L_\alpha = \mathbb{1} \otimes \tilde{L}_\alpha$ with $\tilde{L}_\alpha = \gamma(|\alpha\rangle_t \langle \alpha+1| + |\alpha+1\rangle_t \langle \alpha|)$.

This unitary transformation renders the convergence rate of the Liouvillian independent of the actual nature of the computation. Without loss of generality, we therefore assume from now on that $\forall \alpha : U_\alpha = \mathbb{1}$. The only two parameters which determine the speed of the computation are the number of two qubit gates $(M + 1)$ in the circuit, and the number of logical qubits involved in the computation (N) .

The critical step in proving rapid convergence of this process is to perform a similarity transformation which casts the map $e^{t\hat{\mathcal{L}}}$ into block diagonal form. The similarity transform considered is $Y = \mathbb{1}_{qq} \otimes \mathbb{1}_{tt} + (X^{\otimes N} - \mathbb{1}_{qq}) \otimes (\sum_{\alpha=0}^M |\alpha\alpha\rangle_t \langle \alpha\alpha|)$, where

$$X = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad (5.11)$$

It can easily be checked that $Y^{-1} = Y$. We point out also that the ordering of the indices has been changed here. Indeed, the Hilbert space ordering is now $\mathcal{H}_q \otimes \mathcal{H}_q \otimes \mathcal{H}_t \otimes \mathcal{H}_t$, where the doubling of logical (q) and time (t) degrees of freedom is a consequence of the linear operator representation of the map. Applying this similarity transformation to $\hat{\mathcal{L}}$, one obtains (after a slightly tedious calculation): $\tilde{\mathcal{L}} \equiv Y\hat{\mathcal{L}}Y = \tilde{\mathcal{L}}_{\text{init}} + \tilde{\mathcal{L}}_{\text{comp}}$ with

$$\tilde{\mathcal{L}}_{\text{init}} = -\frac{1}{2}\gamma \sum_i (|1\rangle_i \langle 1| \otimes \mathbb{1}_q \otimes |0\rangle_t \langle 0| \otimes \mathbb{1}_t + \mathbb{1}_q \otimes |1\rangle_i \langle 1| \otimes \mathbb{1}_t \otimes |0\rangle_t \langle 0|) \quad (5.12)$$

$$\tilde{\mathcal{L}}_{\text{comp}} = \mathbb{1}_{qq} \otimes \tilde{\mathcal{L}} \equiv \sum_{\alpha} \mathbb{1}_{qq} \otimes (\tilde{\mathcal{L}}_{\alpha} \otimes \tilde{\mathcal{L}}_{\alpha} - \frac{1}{2}(\tilde{\mathcal{L}}_{\alpha}^{\dagger} \tilde{\mathcal{L}}_{\alpha} \otimes \mathbb{1}_t + \mathbb{1}_t \otimes \tilde{\mathcal{L}}_{\alpha}^{\dagger} \tilde{\mathcal{L}}_{\alpha})) \quad (5.13)$$

Clearly, $\tilde{\mathcal{L}}$ is diagonal in the logical basis, and its block diagonal elements can be rewritten explicitly as

$$L_{ij} = \tilde{\mathcal{L}} - \frac{1}{2}(\gamma_i |0\rangle_t \langle 0| \otimes \mathbb{1}_t + \gamma_j \mathbb{1}_t \otimes |0\rangle_t \langle 0|), \quad (5.14)$$

where $\gamma_i = \gamma \sum_{k=1}^N i_k$ is the number of ones in the binomial expansion of the block diagonal index. Each block can then be further decomposed into: one dimensional blocks with values (-2γ) , $-\frac{1}{2}(3\gamma + \gamma_i)$, $-\frac{1}{2}(3\gamma + \gamma_j)$; two dimensional blocks of the form

$$-\frac{1}{2} \begin{pmatrix} 3\gamma + \gamma_i & -2\gamma \\ -2\gamma & 3\gamma + \gamma_j \end{pmatrix}, \gamma \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} \quad (5.15)$$

and tridiagonal matrices of the form $(\gamma E - \frac{1}{2}(\gamma_i + \gamma_j)V)$, where

$$E = \begin{pmatrix} -1 & 1 & & \\ 1 & -2 & 1 & \\ & 1 & -2 & 1 \\ & & \dots & \end{pmatrix}, \quad V = \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & 0 & \\ & & & \dots \end{pmatrix} \quad (5.16)$$

E is the generator of a stochastic matrix, which has been extensively studied in the context of Markov chains, and is often called a *homogeneous birth and death chain*. Its eigenvalues, eigenvectors, and convergence rate can be calculated analytically [DLP08].

We now use some of the contraction properties from the first section along with some basic bounds on the eigenvalues of Markov chains in order to get upper bounds on the convergence of this semigroup. Properties of the operator norm tell us that it is sufficient to bound each of the block diagonal elements separately. By inspection, it is clear that the one- and two-dimensional blocks have spectra whose real part is bounded above by $-\frac{1}{2}$. As the dimension of the blocks and the gap are independent of (N, M) , and the multiplicity of the blocks does not affect the convergence time, it follows that these blocks do not contribute to the convergence for increasing (N, M) . Therefore, we can restrict our attention to the convergence of the tridiagonal blocks.

The effective dimension of each tridiagonal block is $(M + 1)$, so that directly invoking Prop. 20, and the comments following it, we see that in order to prove polynomial convergence time, it is sufficient to bound the gap of the process. In particular, the prefactor can only contribute an order $\mathcal{O}(M^4)$ to the mixing time. Below, we improve this bound as well as bounding the gap.

The spectrum of E is given by $\lambda_k = 2(\cos[\pi k/(M + 1)] - 1)$, ($k = 0, 1, \dots, M$). Clearly, the eigenvalues of E are all non-degenerate, and 0 is an eigenvalue. We therefore get

$$\|e^{tE} - \hat{T}_\infty\| = \left\| \sum_{k=0}^M e^{t\lambda_k} |\phi_k\rangle\langle\phi_k| - |\phi_0\rangle\langle\phi_0| \right\| \quad (5.17)$$

$$\leq \sum_{k=1}^M e^{t\lambda_k} \leq M e^{t\lambda_1}, \quad (5.18)$$

which leads to a mixing time for E of $\mathcal{O}(M^2 \log M)$. The convergence time of this Markov chain actually is $\mathcal{O}(M^2)$, but this improvement is irrelevant to us, as we anyhow picked up a $\log M$ term by bounding the contraction coefficient by the operator norm in Eqn. (5.7). Thus, we have proved that the tridiagonal block with $\gamma_i = \gamma_j = 0$ converges in a time which scales as $\mathcal{O}(NM^2 \log M)$. What about the other blocks?

First note that adding $-\frac{1}{2}(\gamma_i + \gamma_j)V$ to γE will decrease each eigenvalue individually. Therefore, the blocks with $\gamma_i + \gamma_j = 1$ have the longest convergence times of any block, except possibly the one with $\gamma_i = \gamma_j = 0$. In order to compare these two cases, we only need to compare their gaps, as the prefactor is determined by the multiplicities in the rest of the spectrum, which do not occur in our case. We use a Lemma from [KSV00]:

Lemma 51 (Kitaev). *Let $A_1, A_2 \geq 0$, and let $\mathcal{N}_1, \mathcal{N}_2$ be their null spaces, where $\mathcal{N}_1 \cap \mathcal{N}_2 = \{0\}$. Let Δ be the minimum of the gaps of A_1, A_2 , then*

$$A_1 + A_2 \geq 2\Delta \sin^2 \frac{\theta}{2}, \quad (5.19)$$

where θ is the angle between \mathcal{N}_1 and \mathcal{N}_2 .

We can estimate the square of the cosine of the angle θ between the null spaces of E and M explicitly:

$$\cos^2 \theta = \max_{|\psi\rangle \in \mathcal{N}_1} \langle \psi | P_{\mathcal{N}_2} | \psi \rangle \quad (5.20)$$

where $P_{\mathcal{N}_2}$ is the projector onto \mathcal{N}_2 . This can easily be seen to be bounded as $\cos^2 \theta \leq 1 - 1/M$. Thus, the gap of $(\gamma E - \frac{1}{2}(\gamma_i + \gamma_j)V)$ is of order $\mathcal{O}(M^{-3})$. Hence, putting all

other elements together, we have shown that the Master equation defined by Eqns. (5.4) and (5.5) converges in a time of order $\mathcal{O}(NM^3 \log M)$.

5.2.2 Unary encoding of the time register, i.e DQC on $\mathcal{H}_2^{\otimes(N+M)}$

Having shown that the quantum Markov semigroup defined by Eqns. (5.4), (5.5) converges to its fixed point in poly-times, we now move on to showing that it is possible to encode the $(M + 1)$ states of the time register on M qubits in such a way that the convergence remains rapid. The encoding is the simple mapping which identifies an integer $t = 0, 1, \dots, M$ with the state $|1\dots 10\dots 0\rangle$ with t consecutive ones followed by $(M - t)$ zeros. The all-zeros state corresponds to the $|0\rangle_t$ state of the original time register. We denote by \mathcal{Q} the space spanned by the $(M + 1)$ vectors of the above encoding and \mathcal{Q}^\perp its $(2^M - M - 1)$ -dimensional complement. States in \mathcal{Q} will be referred to as *allowed*, while state in \mathcal{Q}^\perp will be referred to as *forbidden*.

The operators in this unary encoding are transformed in the following way:

$$\begin{aligned} |0\rangle_t \langle 0| &\rightarrow |0\rangle_1 \langle 0|, & |0\rangle_t \langle 1| &\rightarrow |0\rangle_1 \langle 1| \otimes |0\rangle_2 \langle 0| \\ |\alpha\rangle_t \langle \alpha| &\rightarrow |1\rangle_\alpha \langle 1| \otimes |0\rangle_{\alpha+1} \langle 0| & |\alpha\rangle_t \langle \alpha + 1| &\rightarrow |1\rangle_{\alpha-1} \langle 1| \otimes |0\rangle_\alpha \langle 1| \otimes |0\rangle_{\alpha+1} \langle 0| \\ |M\rangle_t \langle M| &\rightarrow |1\rangle_M \langle 1|, & |M-1\rangle_t \langle M| &\rightarrow |1\rangle_{M-1} \langle 1| \otimes |0\rangle_M \langle 1| \end{aligned} \quad (5.21)$$

We can then rewrite the Lindblad operators in their unary encodings:

$$L_i^u = \gamma |0\rangle_i \langle 1| \otimes |0\rangle_1 \langle 0| \quad (5.22)$$

$$L_\alpha^u = 1_q \otimes \tilde{L}_\alpha^u \quad (5.23)$$

where \tilde{L}_α^u is the unary encoding of \tilde{L}_α from the transformations in Eqn. (5.21), and the tensor product refers to the new partition $\mathcal{H}_2^{\otimes N} \otimes \mathcal{H}_2^{\otimes M}$. As the transformation is surjective, the Lindblad operators act on \mathcal{Q} in exactly the same manner as do Eqns. (5.4) and (5.5), on the Kitaev clock Hilbert space.

In order to guarantee the uniqueness of the fixed point, however, we need to add another set of Lindblad operators which act trivially on \mathcal{Q} , but "drain" \mathcal{Q}^\perp of all of its elements. These operators are:

$$L_\alpha^{\text{forb}} = \gamma 1_q \otimes (|0\rangle_{\alpha-1} \langle 0| \otimes |0\rangle_\alpha \langle 1|), \quad (5.24)$$

for $\alpha = 1, \dots, M$. It is easy to check, using the Davies-Frigiero-Spohn criterion (Prop. 11), that the semigroup defined by Eqns. (5.22), (5.23), and Eqn. (5.24) also has a unique stationary state which is simply the unary encoding of Eqn. (5.6). We also apply rapid local dephasing noise on the qubits of the unary time register so that the timer states are separable throughout the computation.

We point out here, that unlike in the previous section, where considering more than just the gap for convergence was a question of cosmetics, as it only eliminated a factor of $\mathcal{O}(M^2)$, the naive bound on the pre-factor from Prop. 20 corresponds to a convergence time which is indeed exponential in M . Hence, bounding the pre-factor is crucial for proving rapid convergence of the computation using the unary time register.

In order to be able to use some of the results and methods of the previous chapters, we need to control the new set of Lindblad operators $\{L_k^{\text{forb}}\}$. In particular, it can be seen that naively applying the unary encoding of the similarity transformation (Y^u) does not even bring the logical subspace of $\{L_k^{\text{forb}}\}$ into block diagonal form. The important observation here is that by splitting the unary time register into (Q) and (Q^\perp) , we can actually bound the convergence on the two subspaces separately. Note that:

- $\mathcal{L}_{\text{init}}^u + \mathcal{L}_{\text{comp}}^u$ acts trivially on Q^\perp
- $\mathcal{L}^{\text{forb}}$ acts trivially on Q
- $\mathcal{L}^{\text{forb}}$ strictly depletes Q^\perp
- $\mathcal{L}_{\text{init}}^u + \mathcal{L}_{\text{comp}}^u$ acts on Q in exactly the same way as the simple (non unary encoding) Liouvillian of the previous section.

We can exploit this structure in order to eliminate the contribution of $\mathcal{L}^{\text{forb}}$ in the convergence. First observe that $\mathcal{L}^{\text{forb}}$ causes a transition in a forbidden state on average once every "unit of time" $\tau \equiv 1/\gamma$. Clearly, the (separable) forbidden state which requires the most transitions (called "jumps" in quantum optics) before entering the subspace of allowed states is $|01\dots 1\rangle$. The number of transitions necessary is $(M - 1)$. In other words, we should expect the forbidden subspace to be essentially completely depleted in a time $\mathcal{O}(M)$. We now make this argument rigorous.

We first bound the convergence of the full Liouvillian by the convergence of its parts. In order to do so, we consider the Trotter expansion for $e^{t(\hat{\mathcal{L}}_{\text{init}}^u + \hat{\mathcal{L}}_{\text{comp}}^u) + t\hat{\mathcal{L}}^{\text{forb}}}$. By introducing the shorthand notation: $\hat{T}_a = e^{(\hat{\mathcal{L}}_{\text{init}}^u + \hat{\mathcal{L}}_{\text{comp}}^u)}$ and $\hat{T}_f = e^{\hat{\mathcal{L}}^{\text{forb}}}$, the Trotter formula reads:

$$\lim_{m \rightarrow \infty} \|e^{t\hat{\mathcal{L}}} - (\hat{T}_a^{t/m} \hat{T}_f^{t/m})^m\| = 0. \quad (5.25)$$

Using this, we bound the contraction of the full system by contraction of its parts as follows. Observe that we can wlog. take the supremum over Q^\perp , as otherwise $\mathcal{L}^{\text{perp}}$ would not contribute to the contraction. Then, we show the following Lemma:

Lemma 52. *Let $\mathcal{L} : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a Liouvillian which can be decomposed as $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$. Given the partition $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ of the Hilbert space with $\mathcal{P}_{1,2}$ the projections onto each part, assume that*

- $\mathcal{L}_1(\rho_2) = 0$ for all $\rho_2 \in \mathcal{B}(\mathcal{H}_2)$
- $\mathcal{L}_2(\rho_1) = 0$ for all $\rho_1 \in \mathcal{B}(\mathcal{H}_1)$
- $e^{t\mathcal{L}_1(\rho_1)} \in \mathcal{B}(\mathcal{H}_1)$ for all $\rho_1 \in \mathcal{B}(\mathcal{H}_1)$

Then,

$$(e^{t\mathcal{L}_1} e^{t\mathcal{L}_2})^m \mathcal{P}_2 = \sum_{j=1}^m e^{t(j-1)\mathcal{L}_1} \mathcal{P}_1 (e^{t\mathcal{L}_2} \mathcal{P}_2)^{m-j} + \mathcal{P}_2 (e^{t\mathcal{L}_2} \mathcal{P}_2)^m \quad (5.26)$$

Proof. Note that for any $\rho_2 \in \mathcal{B}(\mathcal{H}_2)$, $e^{t\mathcal{L}_1}(\rho_2) = \rho_2$, and similarly, for any $\rho_1 \in \mathcal{B}(\mathcal{H}_1)$, $e^{t\mathcal{L}_2}(\rho_1) = \rho_1$. Note also that for $\rho_2 \in \mathcal{B}(\mathcal{H}_2)$, $e^{t\mathcal{L}_2}(\rho_2) = \mathcal{P}_1 e^{t\mathcal{L}_2}(\rho_2) + \mathcal{P}_2 e^{t\mathcal{L}_2}(\rho_2)$. Now, using these facts, along with the observation that $\mathcal{P}_1 \mathcal{P}_2 = \mathcal{P}_2 \mathcal{P}_1 = 0$, one can evaluate Eqn. (5.26) explicitly. \square

It follows that for large enough m , and any $s \in (0, 1]$,

$$\begin{aligned}
\sup_{\phi \in \mathcal{Q}^\perp} \|e^{t\mathcal{L}}(\phi) - \rho_{ss}\|_1 &\cong \sup_{\phi \in \mathcal{Q}^\perp} \|(T_a^{t/m} T_f^{t/m})^m(\phi) - \rho_{ss}\|_1 \\
&= \sup_{\phi \in \mathcal{Q}^\perp} \|(T_a^{t/m} T_f^{t/m})^{m(1-s)} \left(\sum_{k=1}^{ms} T_a^{kt/m} \mathcal{P}(T_f^{t/m} \mathcal{P}_\perp)^{ms-k} \right. \\
&\quad \left. + \mathcal{P}_\perp (T_f^{t/m} \mathcal{P}_\perp)^{ms} \right)(\phi) - \rho_{ss}\|_1 \\
&\leq \sup_{\phi \in \mathcal{Q}^\perp} \|(T_a^{t/m} T_f^{t/m})^{m(1-s)} \sum_{k=1}^{ms} T_a^{kt/m} \mathcal{P}(T_f^{t/m} \mathcal{P}_\perp)^{ms-k}(\phi) - \rho_{ss}\|_1 \\
&\quad + \sup_{\phi \in \mathcal{Q}^\perp} \|(T_a^{t/m} T_f^{t/m})^{m(1-s)} \mathcal{P}_\perp (T_f^{t/m} \mathcal{P}_\perp)^{ms}(\phi)\|_1 \\
&\leq \sup_{\psi \in \mathcal{Q}} \|T_a^{t(1-s)}(\psi) - \rho_{ss}\|_1 + \sup_{\phi \in \mathcal{Q}^\perp} \|\mathcal{P}_\perp T_f^{st} \mathcal{P}_\perp(\phi)\|_1
\end{aligned}$$

The last inequality is obtained by monotonicity of the trace distance, and by noting that $\sum_{k=1}^{ms} T_a^{kt/m} \mathcal{P}(T_f^{t/m} \mathcal{P}_\perp)^{ms-k}(\phi)$ has support in \mathcal{Q} . Hence, we have expressed the convergence behavior of \mathcal{L}^u in terms of the convergence of $(\mathcal{L}_{\text{init}}^u + \mathcal{L}_{\text{comp}}^u)$ and $\mathcal{L}^{\text{forb}}$ separately.

We now bound $\sup_{\phi \in \mathcal{Q}^\perp} \|\mathcal{P}_\perp e^{t\mathcal{L}^{\text{forb}}} \mathcal{P}_\perp(\phi)\|_1$. Note first that as $t \rightarrow \infty$, this term vanishes, as the forbidden subspace is strictly depleted by $\mathcal{L}^{\text{forb}}$. Secondly, observe that $\mathcal{L}^{\text{forb}}$ takes diagonal density matrices to diagonal density matrices (in the logical basis), so that the process is a classical continuous time Markov Chain. Denote the generator of the classical Markov chain A so that $A_{ij} \equiv \langle \vec{i} | \mathcal{L}^{\text{forb}}(|\vec{j}\rangle\langle\vec{j}|) | \vec{i} \rangle$ and denote the classical logical basis vectors $v_{\vec{i}}$, where \vec{i} is shorthand notation for $\{i_1, \dots, i_M\}$. A acts as $Av_{i_1, \dots, 0_j, 1_{j+1}, \dots, i_M} = \gamma(v_{i_1, \dots, 0_j, 0_{j+1}, \dots, i_M} - v_{i_1, \dots, 0_j, 1_{j+1}, \dots, i_M})$ for vectors in the forbidden set of states, and it acts trivially on the allowed states. Hence, the entries of A can be re-ordered to give

$$E = \begin{pmatrix} 0 & r \\ 0 & Q \end{pmatrix} \quad (5.27)$$

where the top left entry is an $(M+1) \times (M+1)$ matrix of zeros corresponding to the allowed states, r is an $(M+1) \times (2^M - M - 1)$ matrix representing transitions from the forbidden to the allowed states, and Q is a block diagonal matrix where each block is in Jordan normal form:

$$Q_k = \gamma \begin{pmatrix} -1 & 1 & & \\ & -1 & 1 & \\ & & -1 & 1 \\ & & & \dots \end{pmatrix} \quad (5.28)$$

The dimension of the largest Jordan block is $M-1$ and its principal associated eigenvector is $v_{0,1,\dots,1}$. Given that the time register is subjected to strong dephasing noise, we can

assume that the maximum over $\phi \in \mathcal{S}_d^+$ is restricted to pure states which are diagonal in the timer basis. It follows that

$$\sup_{\phi} \|\mathcal{P}_{\perp} e^{t\mathcal{L}^{\text{form}}} \mathcal{P}_{\perp}(\phi)\|_1 = \sup_v \|e^{tQ}v\|_1 = \max_k \sup_v \|e^{tQ_k}v\|_1 \quad (5.29)$$

The maximum over k is reached for the largest Jordan Block, and the supremum over v is reached for $v_{0,1,\dots,1}$. A simple calculation then yields

$$\max_k \sup_v \|e^{tQ_k}v\|_1 \leq e^{-t\gamma} \sum_{k=0}^{M-1} \frac{(t\gamma)^k}{k!} = \frac{\Gamma(M, t\gamma)}{(M-1)!} \quad (5.30)$$

where $\Gamma(x, y)$ is the incomplete gamma function. It can be shown [Pag65] that $\Gamma(M, cM)/(M-1)! \rightarrow 0$ as $M \rightarrow \infty$ for any $c > 1$.

Thus, we have shown that after a time of order $\mathcal{O}(M)$, it is sufficient to consider the dynamics restricted to \mathcal{Q} . We can then just invoke the surjective unary mapping in order to apply the results of the previous section! Hence, also in the unary encoding, the convergence time of the semigroup is of order $\mathcal{O}(NM^3 \log M)$.

The necessary Lindblad operators are 5-local because they are composed of 2-local unitaries and 3-local operators for the unary encoding of the time register.

Hilbert's projective metric

Synopsis:

We introduce and apply Hilbert's projective metric in the context of quantum information theory. The metric is induced by convex cones such as the sets of positive, separable or PPT operators. It provides bounds on measures for statistical distinguishability of quantum states and on the decrease of entanglement under LOCC protocols or other cone-preserving operations. The results are formulated in terms of general cones and base norms and lead to contractivity bounds for quantum channels, for instance improving Ruskai's trace-norm contraction inequality. A new duality between distinguishability measures and base norms is provided. For two given pairs of quantum states we show that the contraction of Hilbert's projective metric is necessary and sufficient for the existence of a probabilistic quantum operation that maps one pair onto the other. Inequalities between Hilbert's projective metric and the Chernoff bound, the fidelity and various norms are proven.

Attached

Hilbert's projective metric in quantum information theory

D. Reeb, M. J. Kastoryano, M. M. Wolf

J. Math. Phys. **52**, 082201 (2011).

Hilbert's projective metric in quantum information theory

David Reeb,^{a)} Michael J. Kastoryano,^{b)} and Michael M. Wolf^{c)}Department of Mathematics, Technische Universität München, 85748 Garching, Germany
and Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark

(Received 26 March 2011; accepted 1 July 2011; published online 4 August 2011)

We introduce and apply Hilbert's projective metric in the context of quantum information theory. The metric is induced by convex cones such as the sets of positive, separable or positive partial transpose operators. It provides bounds on measures for statistical distinguishability of quantum states and on the decrease of entanglement under protocols involving local quantum operations and classical communication or under other cone-preserving operations. The results are formulated in terms of general cones and base norms and lead to contractivity bounds for quantum channels, for instance, improving Ruskai's trace-norm contraction inequality. A new duality between distinguishability measures and base norms is provided. For two given pairs of quantum states we show that the contraction of Hilbert's projective metric is necessary and sufficient for the existence of a probabilistic quantum operation that maps one pair onto the other. Inequalities between Hilbert's projective metric and the Chernoff bound, the fidelity and various norms are proven. © 2011 American Institute of Physics. [doi:10.1063/1.3615729]

I. INTRODUCTION

Convex cones lurk around many corners in quantum information theory—examples include the set of positive semidefinite operators or the subset of separable operators, i.e., the cone generated by unentangled density matrices. These cones come together with important classes of cone-preserving linear maps, such as quantum channels, which preserve positivity, or local operations with classical communication (LOCC maps), which in addition preserve the cone of separable operators.

In the present work, we investigate a distance measure that naturally arises in the context of cones and cone-preserving maps—*Hilbert's projective metric*—from the perspective of quantum information theory. The obtained results are mostly formulated in terms of general cones and subsequently reduced to special cases for the purposes of quantum information theory. Our findings come in two related flavors: (i) inequalities between Hilbert's projective metric and other distance measures (Secs. III, V, and VI), and (ii) contraction bounds for cone-preserving maps (Secs. IV, V, and VII). The latter follows the spirit of Birkhoff's work⁷ in which Hilbert's projective metric was used to prove and extend results of Perron-Frobenius theory (see especially Theorem 4 below).

Before going into detail, we sketch and motivate some of the main results of our work in quantum information terms (for which we refer to Ref. 28 for an introduction):

- **Contraction bounds.** A basic inequality in quantum information theory states that the trace-norm distance of two quantum states ρ_1, ρ_2 is never increased by the application of a quantum channel T , i.e.,

$$\|T(\rho_1) - T(\rho_2)\|_1 \leq \eta \|\rho_1 - \rho_2\|_1, \quad (1)$$

with $\eta = 1$.³³ In Sec. IV, we will generalize this inequality to arbitrary *base norms* and sharpen it in Corollary 9 by some $\eta \leq 1$ that depends on the *diameter* of the image of T when measured in terms of Hilbert's projective metric; see Eq. (38).

^{a)}Electronic mail: david.reeb@tum.de.

^{b)}Electronic mail: kastoryano@nbi.dk.

^{c)}Electronic mail: m.wolf@tum.de.

- **Bounds on distinguishability measures.** The operational meaning of the trace-norm distance, which appears in Eq. (1), is that of a measure of statistical distinguishability when arbitrary measurements are allowed for. If the set M of measurements is restricted, e.g., to those implementable by LOCC operations, the relevant distance measure is given by a different norm:²⁷

$$\|\rho_1 - \rho_2\|_{(M)} = \sup_{E \in M} \text{tr} [2E(\rho_1 - \rho_2)]. \quad (2)$$

Section V shows how such norms can be bounded in terms of Hilbert's projective metric. These results are based on a duality between *distinguishability norms* (2) and base norms; see Theorem 14, from which also a contraction result (Proposition 16) for general distinguishability measures will follow.

- **Bounds on other distance measures in quantum information theory.** In a similar vein, Hilbert's projective metric between two quantum states also bounds their *fidelity* (Proposition 18) and the *Chernoff bound* that quantifies their asymptotic distinguishability in symmetric hypothesis testing (Proposition 20).
- **Decrease of entanglement.** If LOCC operation maps $\rho \mapsto \rho_i$ with probability p_i , then

$$\sum_i p_i \mathcal{N}(\rho_i) \leq \eta \mathcal{N}(\rho), \quad (3)$$

with $\eta = 1$ and \mathcal{N} denoting a *negativity* which quantifies entanglement.³⁷ This means that entanglement is on average non-increasing under LOCC operations, i.e., \mathcal{N} is an *entanglement monotone*.¹⁹ In Proposition 13 we show that $\eta \leq 1$ can be specified in terms of Hilbert's projective metric.

- **Partially specified quantum operations.** In Sec. VII we show that, given two pairs of quantum states, the mapping $\rho_i \mapsto \rho'_i$ with $i = 1, 2$ can be realized probabilistically by a single quantum operation, i.e., $T(\rho_i) = p_i \rho'_i$ for some $p_i > 0$, if and only if their distance with respect to Hilbert's projective metric is non-increasing; see Theorem 21 and subsequent discussion.

The outline of the paper is as follows. In Sec. II we define Hilbert's projective metric, illustrate it in the context of quantum information theory, and summarize the classic results related to it. In Sec. III we connect Hilbert's projective metric to base norms and negativities, quantities that are frequently used in entanglement theory and in other areas of quantum information theory and whose definition is based on cones as well. In Sec. IV we turn to dynamics and consider linear maps whose action preserves cones. We prove that, under the action of such cone-preserving maps, base norms and negativities contract by non-trivial factors that can be expressed via Hilbert's projective metric. In Sec. V we define general norms that arise as distinguishability measures in the quantum information context and illustrate them by physical examples (e.g., measuring the LOCC distinguishability of two quantum states). Via a new duality theorem, we relate these distinguishability norms to the aforementioned base norms. We are thus able to connect the distinguishability norms and their contractivity properties to Hilbert's projective metric. In Sec. VI we prove upper bounds on the quantum fidelity and on the quantum Chernoff bound in terms of Hilbert's projective metric. For the special case of the positive semidefinite cone, we present an operational interpretation of Hilbert's projective metric in Sec. VII as the criterion deciding the physical implementability of a certain operation on given quantum states. We conclude in Sec. VIII.

As examples, in Appendices A and B we consider Hilbert's projective metric for qubits and in the context of depolarizing channels, illustrating results from the main text. As is reflected in these examples, Hilbert's projective metric and the projective diameter seem to be hard to compute exactly in many situations, but nonetheless they can serve as theoretical tools, for instance, guaranteeing non-trivial contraction factors that are otherwise hard to obtain. In Appendix C we show the optimality of several of the bounds from the main text of the paper.

In Secs. II–V we develop the formalism and prove the statements first for general cones and bases. Interspersed into this exposition are then paragraphs and examples which translate the general framework explicitly to the context of quantum information theory, and these paragraphs appear with their initial words in bold font for quick accessibility.

II. BASIC CONCEPTS

In this section we will recall some basic notions from convex analysis and summarize some of the main definitions and results related to Hilbert's projective metric (see Refs. 8, 9, and 14). Throughout we will consider finite-dimensional real vector spaces which we denote by \mathcal{V} . We will mostly think of \mathcal{V} as the space of Hermitian matrices in $\mathcal{M}_d(\mathbb{C})$, in which case $\mathcal{V} \simeq \mathbb{R}^{d^2}$ and there is a standard choice of inner product $\langle a, b \rangle = \text{tr}[ab]$ for $a, b \in \mathcal{V}$; see the quantum theory example later in this section. A *convex cone* $\mathcal{C} \subset \mathcal{V}$ is a subset for which $\alpha\mathcal{C} + \beta\mathcal{C} \subseteq \mathcal{C}$ for all $\alpha, \beta \geq 0$. We will call a convex cone *pointed* if $\mathcal{C} \cap (-\mathcal{C}) = \{0\}$, and *solid* if $\text{span } \mathcal{C} = \mathcal{V}$ (For a basic reference on convex analysis, see also Ref. 32).³⁹ The *dual cone* defined as $\mathcal{C}^* := \{v \in \mathcal{V}^* | \forall c \in \mathcal{C} : \langle v, c \rangle \geq 0\}$ is closed and convex, and, by the *bipolar theorem*, $\mathcal{C}^{**} = \mathcal{C}$ holds if \mathcal{C} is closed and convex. In this case,

$$c \in \mathcal{C} \Leftrightarrow \langle v, c \rangle \geq 0 \quad \text{for all } v \in \mathcal{C}^*. \quad (4)$$

A closed convex cone is solid if and only if its dual is pointed.

Convex cones which are pointed and closed are in one-to-one correspondence with *partial orders* in \mathcal{V} . We will write $a \geq_{\mathcal{C}} b$ meaning $a - b \in \mathcal{C}$, and if determined by the context we will omit the subscript \mathcal{C} . If, for instance, $\mathcal{C} = \mathcal{S}_+$ is the cone of positive semidefinite matrices, then $a \geq b$ is the usual operator ordering. Consistent with this example and having the partial order in mind, one often refers to the elements of the cone as the *positive* elements of the vector space.

For the sake of brevity we will call \mathcal{C} a *proper cone* if it is a closed, convex, pointed, and solid cone within a finite-dimensional real vector space. A convex set $\mathcal{B} \subset \mathcal{V}$ is said to *generate* the convex cone \mathcal{C} if $\mathcal{C} = \bigcup_{\lambda \geq 0} \lambda \mathcal{B}$. Convex sets that are of interest in quantum information theory²⁸ are, for example, those of density matrices, separable states, positive partial transpose (PPT) states, PPT operators, effect operators of POVMs (positive operator valued measures), effect operators reachable via LOCC or PPT operations, etc. As all of these sets generate proper cones, we will in the following focus on proper cones \mathcal{C} . Note that the dual cone \mathcal{C}^* is then a proper cone as well.

For every pair of non-zero elements $a, b \in \mathcal{C}$ define

$$\sup(a/b) := \sup_{v \in \mathcal{C}^*} \frac{\langle v, a \rangle}{\langle v, b \rangle}, \quad \inf(a/b) := \inf_{v \in \mathcal{C}^*} \frac{\langle v, a \rangle}{\langle v, b \rangle}, \quad (5)$$

with the extrema taken over all $v \in \mathcal{C}^*$ leading to a non-zero denominator. By construction, $\sup(a/b) = 1/\inf(b/a)$ and $\sup(a/b) \geq \inf(a/b) \geq 0$. Their difference was studied by Hopf²² and is called *oscillation* $\text{osc}(a/b) := \sup(a/b) - \inf(a/b)$. The oscillation is invariant under the substitution $a \rightarrow a + \beta b$ for any $\beta \in \mathbb{R}$.

If \mathcal{C} is a proper cone, we can use Eq. (4) to rewrite

$$\sup(a/b) = \inf\{\lambda \in \mathbb{R} | a \leq_{\mathcal{C}} \lambda b\}, \quad (6)$$

$$\inf(a/b) = \sup\{\lambda \in \mathbb{R} | \lambda b \leq_{\mathcal{C}} a\}, \quad (7)$$

with the convention that $\sup(a/b) = \infty$ if there is no λ such that $a \leq_{\mathcal{C}} \lambda b$. This implies that $\inf(a/b)b \leq_{\mathcal{C}} a \leq_{\mathcal{C}} \sup(a/b)b$, where the last inequality makes sense only if $\sup(a/b)$ is finite. In other words, Eq. (5) provides the factors by which b has to be rescaled at least in order to become larger or smaller than a .

Hilbert's projective metric is defined for $a, b \in \mathcal{C} \setminus 0$ as^{8,14}

$$\mathfrak{h}(a, b) := \ln [\sup(a/b) \sup(b/a)], \quad (8)$$

and one defines $\mathfrak{h}(0, 0) := 0$ and $\mathfrak{h}(0, a) := \mathfrak{h}(a, 0) := \infty$ for $a \in \mathcal{C} \setminus 0$. Keep in mind that \mathfrak{h} , \sup , \inf , and osc all depend on the chosen cone \mathcal{C} which we will thus occasionally use as a subscript and, for instance, write $\mathfrak{h}_{\mathcal{C}}$ if confusion is ahead.⁴⁰ Obviously, \mathfrak{h} is symmetric, non-negative and satisfies $\mathfrak{h}(a, \beta b) = \mathfrak{h}(a, b)$ for all $\beta > 0$. That is, \mathfrak{h} depends only on the “direction” of its arguments. Since it satisfies the triangle inequality (due to $\sup(a/b) \sup(b/c) \geq \sup(a/c)$) and since $\mathfrak{h}(a, b) = 0$ implies that $a = \beta b$ for some $\beta > 0$, \mathfrak{h} is a *projective metric* on \mathcal{C} . Hence, if we restrict the arguments a, b further to a subset which excludes multiples of elements (e.g., to the unit sphere of a norm, or to a hyperplane that contains a set generating the cone), then \mathfrak{h} becomes a metric on that space. Note that

Hilbert's projective metric puts any boundary point of the cone at infinite distance from every interior point,⁴¹ whereas two interior points always have finite distance. As for distances induced by norms, Hilbert's projective metric is additive on lines, $\mathfrak{h}(a, b) + \mathfrak{h}(b, c) = \mathfrak{h}(a, c)$ for $b = \lambda a + (1 - \lambda)c$ with $0 \leq \lambda \leq 1$.

Paradigmatic applications. As alluded to above, when taking $\mathcal{V} \simeq \mathbb{R}^{d^2}$ to be the real vector space of Hermitian matrices in $\mathcal{M}_d(\mathbb{C})$, the cone $\mathcal{S}_+ \subset \mathcal{V}$ of positive semidefinite matrices is proper and contains all density matrices on a d -dimensional quantum system. In fact, the set \mathcal{B}_+ of all density matrices is the intersection of \mathcal{S}_+ with the hyperplane of *normalized* matrices (i.e., those with trace 1), so \mathcal{B}_+ generates \mathcal{S}_+ , and $\mathfrak{h}_{\mathcal{S}_+}$ is a metric on the set of density matrices. In this vector space, there is a standard choice of inner product $\langle a, b \rangle := \text{tr}[ab]$, $a, b \in \mathcal{V}$, so that one has a natural identification $\mathcal{V} \simeq \mathcal{V}^*$ and $(\mathcal{S}_+)^* \simeq \mathcal{S}_+$. Then, in the quantum context, one can give an interpretation to definition (5): for normalized quantum states $\rho, \sigma \in \mathcal{B}_+$ (for which we will often use these Greek letters), $\sup_{\mathcal{S}_+}(\rho/\sigma)$ equals the supremum of $\text{tr}[E\rho]/\text{tr}[E\sigma]$ over all $E \in \mathcal{S}_+$ and is thus the largest possible ratio of probabilities of any measurement outcome (corresponding to E) on the state ρ versus on σ . Furthermore, from expression (6), $\sup_{\mathcal{S}_+}(\rho/\sigma)$ equals—up to a logarithm—the *max-relative entropy* of ρ and σ .¹²

Other convex sets and cones of interest in quantum information theory will be discussed in Secs. III and V. The first classic application of Hilbert's projective metric was to the vector space $\mathcal{V} = \mathbb{R}^d$ with the cone $\mathcal{C} = (\mathbb{R}_+)^d$ of vectors with non-negative entries (un-normalized probability vectors). Perron-Frobenius theory can be developed in this context,⁷ and one can compute for $p, q \in \mathcal{C} \setminus 0$,

$$\sup(p/q) = \max_{1 \leq i \leq d} \frac{p_i}{q_i}, \quad \inf(p/q) = \min_{1 \leq i \leq d} \frac{p_i}{q_i},$$

omitting indices i for which $p_i = q_i = 0$, and defining $p_i/0 := \infty$ for $p_i > 0$. Similarly, for the cone \mathcal{S}_+ of positive semidefinite operators from the previous paragraph, one can explicitly compute all of the above defined quantities so that their properties may become more transparent (cf., also the qubit example in Appendix A):

Proposition 1 (Hilbert distance with respect to positive semidefinite cone): Consider the cone \mathcal{S}_+ of positive semidefinite matrices in $\mathcal{M}_d(\mathbb{C})$ and let $A, B \in \mathcal{S}_+$. Then, with $(\cdot)^{-1}$ denoting the pseudoinverse (inverse on the support) and with $\|\cdot\|_\infty$ being the operator norm, we have

$$\begin{aligned} \sup(A/B) &= \begin{cases} \|B^{-1/2}AB^{-1/2}\|_\infty, & \text{if } \text{supp}[A] \subseteq \text{supp}[B], \\ \infty, & \text{otherwise} \end{cases} \\ \inf(A/B) &= \begin{cases} \|A^{-1/2}BA^{-1/2}\|_\infty^{-1}, & \text{if } \text{supp}[B] \subseteq \text{supp}[A], \\ 0, & \text{otherwise} \end{cases} \\ \mathfrak{h}_{\mathcal{S}_+}(A, B) &= \begin{cases} \ln \left[\|A^{-1/2}BA^{-1/2}\|_\infty \|B^{-1/2}AB^{-1/2}\|_\infty \right], & \text{if } \text{supp}[B] = \text{supp}[A] \\ \infty, & \text{otherwise.} \end{cases} \end{aligned} \quad (9)$$

Proof: We only have to prove the relation for $\sup(A/B)$ since this implies the other two by $\inf(A/B) = 1/\sup(B/A)$ and definition (8), respectively. Assume that $\text{supp}[A] \not\subseteq \text{supp}[B]$. Then there is a vector $\psi \in \mathbb{C}^d$ for which $\langle \psi|B|\psi \rangle = 0$ while $\langle \psi|A|\psi \rangle > 0$, so that the infimum in (6) is over an empty set and thus by definition ∞ . If, however, $\text{supp}[A] \subseteq \text{supp}[B]$, then $A \leq \lambda B$ is equivalent to $B^{-1/2}AB^{-1/2} \leq \lambda I$ and the smallest λ for which this holds is the operator norm. \square

Multiplicativity of the operator norm gives the following:

Corollary 2 (Additivity on tensor products): For $i = 1, 2$, denote by $\mathcal{S}_{(i)+}$ and by \mathcal{S}_+ the cones of positive semidefinite matrices in $\mathcal{M}_{d_i}(\mathbb{C})$ and in $\mathcal{M}_{d_1}(\mathbb{C}) \otimes \mathcal{M}_{d_2}(\mathbb{C})$, respectively, and let $A_i, B_i \in \mathcal{S}_{(i)+}$. Then

$$\mathfrak{h}_{\mathcal{S}_+}(A_1 \otimes A_2, B_1 \otimes B_2) = \mathfrak{h}_{\mathcal{S}_{(1)+}}(A_1, B_1) + \mathfrak{h}_{\mathcal{S}_{(2)+}}(A_2, B_2). \quad (10)$$

Contraction properties for positive maps is the main context in which Hilbert's projective metric is applied. A map $T : \mathcal{V} \rightarrow \mathcal{V}'$ between two partially ordered vector spaces with corresponding cones \mathcal{C} and \mathcal{C}' is called *cone-preserving* or *positive* if it maps one cone (which corresponds to the set of “positive elements”) into the other, i.e., $T(\mathcal{C}) \subseteq \mathcal{C}'$. We will in the following exclusively consider linear maps, although parts of the theory also apply to homogeneous maps of degree smaller than one.⁸ In many cases one has $\mathcal{C} = \mathcal{C}'$, but one can imagine applications where different cones appear: if, for instance, a quantum channel T maps any bipartite density matrix onto a separable one or onto one with a certain support or symmetry, we may choose different cones for input and output.

As an important notion in analyzing contractivity properties, we will need

Definition 3 (Projective diameter): For proper cones $\mathcal{C} \subset \mathcal{V}$ and $\mathcal{C}' \subset \mathcal{V}'$, let $T : \mathcal{C} \rightarrow \mathcal{C}'$ be a positive linear map. Then the projective diameter of the image of T , or, for short, the projective diameter of T , is defined as

$$\Delta(T) := \sup_{a,b \in \mathcal{C} \setminus 0} \mathfrak{h}_{\mathcal{C}'}(T(a), T(b)). \quad (11)$$

The central theorem, which has its origin in Birkhoff's analysis of Perron-Frobenius theory, is the following.^{5,7,14,22}

Theorem 4 (Birkhoff-Hopf contraction theorem): Let $T : \mathcal{C} \rightarrow \mathcal{C}'$ be a positive linear map between two proper cones \mathcal{C} and \mathcal{C}' . Then, denoting by $\mathcal{K} \subset \mathcal{C} \times \mathcal{C}$ the set of pairs (a, b) for which $0 < \mathfrak{h}_{\mathcal{C}}(a, b) < \infty$,

$$\sup_{(a,b) \in \mathcal{K}} \frac{\mathfrak{h}_{\mathcal{C}'}(T(a), T(b))}{\mathfrak{h}_{\mathcal{C}}(a, b)} = \sup_{(a,b) \in \mathcal{K}} \frac{\text{osc}_{\mathcal{C}'}(T(a)/T(b))}{\text{osc}_{\mathcal{C}}(a/b)} = \tanh \frac{\Delta(T)}{4}. \quad (12)$$

In other words, any positive map T is a contraction with respect to Hilbert's projective metric (and the oscillation) and $\eta^{\mathfrak{h}}(T) := \tanh[\Delta(T)/4] \in [0, 1]$ is the *best possible* contraction coefficient. As a consequence we get that this coefficient is sub-multiplicative in the sense that for a composition of positive maps we have

$$\eta^{\mathfrak{h}}(T_2 T_1) \leq \eta^{\mathfrak{h}}(T_2) \eta^{\mathfrak{h}}(T_1).$$

Thus, if $\mathcal{C} = \mathcal{C}'$, then $\eta^{\mathfrak{h}}(T^n) \leq \eta^{\mathfrak{h}}(T)^n$ for all $n \in \mathbb{N}$. Moreover, and this is Birkhoff's observation, if $\Delta(T^m) < \infty$ for some $m \in \mathbb{N}$, then there exists a “fixed point” (or better “fixed ray”) $T(c) \propto c \in \mathcal{C} \setminus 0$ that is unique up to scalar multiplication. The uniqueness of a fixed point, a central statement of Perron-Frobenius theory, is often related to spectral properties of the considered map. The following shows how the above contraction coefficient is related to the spectrum.¹⁴

Theorem 5 (Spectral bound on projective diameter): Let $\mathcal{C} \subset \mathcal{V}$ be a proper cone and $T : \mathcal{C} \rightarrow \mathcal{C}$ a positive linear map with $T(c) = c$ for some non-zero $c \in \mathcal{C}$, and $\Delta(T) < \infty$. If $T(a) = \lambda a$ for some $\lambda \in \mathbb{C}$ and $a \in \mathcal{V} + i\mathcal{V}$ with $a \not\propto c$, then

$$|\lambda| \leq \tanh [\Delta(T)/4]. \quad (13)$$

Consequently, if $\Delta(T^m) < \infty$ for some $m \in \mathbb{N}$ and $T(c) = c \in \mathcal{C} \setminus 0$, then all but one of the eigenvalues of T have modulus strictly smaller than one (even counting algebraic multiplicities¹⁴), so the spectral radius of T equals 1, which is itself an eigenvalue with positive eigenvector.

Having the last two theorems in mind, one may wonder whether there are other constructions of projective metrics that lead to even stronger results. The following shows that Hilbert's approach is in a sense unique and optimal.²³ Stating it requires a general definition of a *projective metric* as a functional $D : \mathcal{C} \times \mathcal{C} \rightarrow \mathbb{R} \cup \infty$ which is non-negative, symmetric, satisfies the triangle inequality, and is such that $D(a, b) = 0$ iff $a = \beta b$ for some positive scalar $\beta > 0$; note that these conditions imply $D(\alpha a, \beta b) = D(a, b)$ for all $a, b \in \mathcal{C}$ and $\alpha, \beta > 0$. Moreover, we call a positive map T a

strict contraction with respect to D , if for all $a, b \in \mathcal{C} \setminus 0$ we have the strict inequality $D(T(a), T(b)) < D(a, b)$ unless $D(a, b) = 0$.

Theorem 6 (Uniqueness of Hilbert's projective metric): *Let \mathcal{C} be a proper cone with interior \mathcal{C}° and let D be a projective metric such that every linear map $T : \mathcal{C} \setminus 0 \rightarrow \mathcal{C}^\circ$ is a strict contraction with respect to D . Then there exists a continuous and strictly increasing function $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that $D(a, b) = f(h_{\mathcal{C}}(a, b))$ for all $a, b \in \mathcal{C}^\circ$, where $h_{\mathcal{C}}$ is Hilbert's projective metric in \mathcal{C} . Moreover, for any linear map $T : \mathcal{C} \rightarrow \mathcal{C}$ we have*

$$\tanh \frac{\Delta(T)}{4} \leq \sup_{a, b \in \mathcal{C} \setminus 0} \left\{ \frac{D(T(a), T(b))}{D(a, b)} \mid D(a, b) > 0 \right\}. \quad (14)$$

As a caveat to the previous theorem, consider the following example. Starting from the trace-norm $\|\cdot\|_1$ on $\mathcal{M}_d(\mathbb{C})$,²⁰ there is an obvious way to define a projective metric D_1 on the cone \mathcal{S}_+ of positive semidefinite matrices ($A, B \in \mathcal{S}_+ \setminus 0$),

$$D_1(A, B) := \left\| \frac{A}{\text{tr}[A]} - \frac{B}{\text{tr}[B]} \right\|_1, \quad D_1(A, 0) := D_1(0, A) := 1, \quad D_1(0, 0) := 0.$$

Theorem 6 does *not* apply to D_1 , as one can find a map $T : \mathcal{S}_+ \setminus 0 \rightarrow (\mathcal{S}_+)^\circ$ and $A, B \in \mathcal{S}_+$ with $D_1(T(A), T(B)) > D_1(A, B)$, so that T is not a strict contraction with respect to D_1 . Importantly, however, due to Ruskai's trace-norm contraction inequality,³³ any physical quantum channel T (i.e., additionally satisfying $\text{tr}[T(A)] = \text{tr}[A]$ for all $A \in \mathcal{S}_+$) is a contraction with respect to D_1 . Moreover, as will be shown later in Corollary 9, inequality (14) is actually reversed in this case, i.e., the contraction coefficient of T with respect to D_1 is *better* (smaller) than with respect to Hilbert's projective metric (see also below Proposition 12). The construction above can more generally be made with base norms, to which we now turn and of which the trace-norm is one example.

III. BASE NORMS AND NEGATIVITIES

In this section, we will first introduce some norms and similar quantities whose definitions are, like Hilbert's projective metric, based on cones, and then show in which guise they appear in quantum information theory, in particular in the theory of entanglement. At the end of this section and in the following one, we will then show how these quantities are related to Hilbert's projective metric. Connections with distinguishability measures in quantum information theory will become apparent in Sec. V.

A *base* \mathcal{B} for a proper cone $\mathcal{C} \subset \mathcal{V}$ is a convex subset $\mathcal{B} \subset \mathcal{C}$ such that every non-zero $c \in \mathcal{C} \setminus 0$ has a unique representation of the form $c = \lambda b$ with $\lambda > 0$ and $b \in \mathcal{B}$. Then \mathcal{B} generates the cone, $\mathcal{C} = \bigcup_{\lambda \geq 0} \lambda \mathcal{B}$, and there exists a unique codimension-1 hyperplane $H_e := \{v \in \mathcal{V} \mid \langle e, v \rangle = 1\}$, corresponding to some linear functional $e \in (\mathcal{C}^*)^\circ$, such that $\mathcal{B} = \mathcal{C} \cap H_e$. Conversely, any compact convex subset \mathcal{B} of a hyperplane that avoids the origin generates a cone \mathcal{C} , which will be proper if and only if the real linear span of \mathcal{B} is all of \mathcal{V} ; the set \mathcal{B} will be a base of \mathcal{C} . Any base of a proper cone equips the vector space \mathcal{V} with a norm, called *base norm*.¹ Introducing the convex hull $\mathcal{B}_\pm := \text{conv}(\mathcal{B} \cup -\mathcal{B})$, the base norm of $v \in \mathcal{V}$ can be defined in several equivalent ways as

$$\|v\|_{\mathcal{B}} := \inf \{ \lambda \geq 0 \mid v \in \lambda \mathcal{B}_\pm \}, \quad (15)$$

$$= \inf \{ \langle e, c_+ \rangle + \langle e, c_- \rangle \mid v = c_+ - c_-, c_\pm \in \mathcal{C} \}, \quad (16)$$

$$= \inf \{ \lambda_+ + \lambda_- \mid v = \lambda_+ b_+ - \lambda_- b_-, \lambda_\pm \geq 0, b_\pm \in \mathcal{B} \}, \quad (17)$$

and \mathcal{B}_\pm will be the unit ball of the base norm $\|\cdot\|_{\mathcal{B}}$. The base norm has the property that $\|v\|_{\mathcal{B}} = \langle e, v \rangle$ if and only if $v \in \mathcal{C}$. In a similar vein, we can define the *negativity*

$$\mathcal{N}_{\mathcal{B}}(v) := \inf \{ \langle e, c_- \rangle \mid v = c_+ - c_-, c_\pm \in \mathcal{C} \}, \quad (18)$$

which is then related to the base norm via $\|v\|_{\mathcal{B}} = \langle e, v \rangle + 2\mathcal{N}(v)$, and which satisfies $\mathcal{N}(v) = 0$ iff $v \in \mathcal{C}$. Somewhat confusingly, in the entanglement theory literature and especially for $v \in H_e$, the quantity $\log \|v\|_{\mathcal{B}}$ is called *logarithmic negativity*⁴² of v .

Paradigmatic application. Continuing the quantum theory example from Sec. II, where \mathcal{V} is the space of Hermitian matrices $A \in \mathcal{M}_d(\mathbb{C})$, by default we take as the base hyperplane H_e the set of *normalized* matrices, $\text{tr}[A] = 1$. In this case the linear functional $e := 1$ is nothing but the trace functional, i.e., $\langle e, A \rangle = \langle 1, A \rangle := \text{tr}[A]$. Using this special functional, the base is determined by specifying the cone, and we can thus employ the usual notation in entanglement theory³⁷ and indicate the base norms $\|\cdot\|_{\mathcal{C}}$ and negativities $\mathcal{N}_{\mathcal{C}}$ by the cone \mathcal{C} rather than by the base \mathcal{B} , which we do in the general case (15) and (18). In quantum theory, all quantum states (density matrices) lie on the hyperplane H_1 . In particular, the set of all *density matrices* $\mathcal{B}_+ := \mathcal{S}_+ \cap H_1$ forms a base for the cone \mathcal{S}_+ of positive semidefinite matrices, and in this case the base norm (16) equals the well-known trace-norm on Hermitian matrices, $\|A\|_{\mathcal{S}_+} = \|A\|_1$ (cf., Ref. 20).

More generally, for any proper cone $\mathcal{C} \subseteq \mathcal{S}_+$, the set $\mathcal{B} := \mathcal{C} \cap H_1 = \{A \in \mathcal{C} | \text{tr}[A] = 1\}$ of quantum states in the cone will be a base for \mathcal{C} . For example, on a bipartite quantum system, where \mathcal{V} is the space of Hermitian matrices in $\mathcal{M}_{d_1 d_2}(\mathbb{C}) \simeq \mathcal{M}_{d_1}(\mathbb{C}) \otimes \mathcal{M}_{d_2}(\mathbb{C})$, the set of *separable matrices*,

$$\mathcal{S}_{\text{SEP}} := \left\{ A \in \mathcal{M}_{d_1 d_2}(\mathbb{C}) \mid A = \sum_k A_k^{(1)} \otimes A_k^{(2)}, A_k^{(i)} \in \mathcal{M}_{d_i}(\mathbb{C}), A_k^{(i)} \text{ positive semidefinite} \right\}, \quad (19)$$

forms a proper cone. This is a subcone of \mathcal{S}_+ , and the set $\mathcal{B}_{\text{SEP}} := \mathcal{S}_{\text{SEP}} \cap H_1$ is a base, the set of all *separable states* on this bipartite system. Even more generally, some cones \mathcal{C} appearing in quantum information theory are not subsets of \mathcal{S}_+ . But whenever the identity matrix is an interior point of the dual cone, i.e., $1 \in (\mathcal{C}^*)^\circ$, one can take the trace functional $\langle 1, A \rangle := \text{tr}[1A]$ to define a base of \mathcal{C} . An example is the cone of matrices with positive partial transpose (*PPT matrices*),

$$\mathcal{S}_{\text{PPT}} := \{A \in \mathcal{M}_{d_1 d_2}(\mathbb{C}) \mid A^{T_1} \text{ positive semidefinite}\} = (\mathcal{S}_+)^{T_1}, \quad (20)$$

where the *partial transposition* T_1 of the first subsystem is defined on tensor products as $(A_1 \otimes A_2)^{T_1} := (A_1)^T \otimes A_2$ and extended to all of $\mathcal{M}_{d_1 d_2}(\mathbb{C})$ by linearity; here, T denotes the usual matrix transposition in $\mathcal{M}_{d_1}(\mathbb{C})$. The cone $\mathcal{S}_{\text{PPT}^+} := \mathcal{S}_+ \cap \mathcal{S}_{\text{PPT}}$ generated by all *PPT states* is another proper cone popular in quantum information theory. Still other cones, for example, generalizations of the above to multipartite quantum systems, can be easily treated in this framework as well.

The base norms and negativities associated to the cones $\mathcal{C} = \mathcal{S}_{\text{PPT}}, \mathcal{S}_{\text{SEP}}, \mathcal{S}_{\text{PPT}^+}$ are used in entanglement theory¹⁹ as measures of entanglement.³⁷ For a normalized bipartite quantum state $\rho \in \mathcal{M}_{d_1 d_2}(\mathbb{C})$, the measures $\mathcal{N}_{\mathcal{C}}(\rho)$ and $\log \|\rho\|_{\mathcal{C}}$ indicate “how far away” a state ρ is from the cone \mathcal{C} , the idea being that all states in those cones \mathcal{C} possess only a weak form of entanglement,^{18,29} or none at all. All of these (generalized) negativities and logarithmic negativities are so-called *entanglement monotones*,^{30,37} see discussion after Proposition 13 below. In particular, the base norm and the negativity corresponding to the cone \mathcal{S}_{PPT} are efficiently computable as $2\mathcal{N}_{\mathcal{S}_{\text{PPT}}}(\rho) + 1 = \|\rho\|_{\mathcal{S}_{\text{PPT}}} = \|\rho\|_{(\mathcal{S}_+)^{T_1}} = \|\rho^{T_1}\|_{\mathcal{S}_+} = \|\rho^{T_1}\|_1$, and in quantum information theory $\mathcal{N}_{\mathcal{S}_{\text{PPT}}}$ is known as *the negativity*. Furthermore, $\mathcal{N}_{\mathcal{S}_{\text{SEP}}}(\rho)$ is usually called *robustness of entanglement*.³⁶

The next proposition relates the distance in base norm between two elements of a cone to their Hilbert distance. It will be used to prove some contractivity results in Secs. IV and V, and a direct interpretation of this proposition from a quantum information perspective will follow from Sec. V, already foreshadowed by the fact that the trace-norm $\|\rho_1 - \rho_2\|_{\mathcal{B}_+} = \|\rho_1 - \rho_2\|_1$ measures the distinguishability between two quantum states $\rho_1, \rho_2 \in \mathcal{B}_+$.

Proposition 7 (Base norm vs. Hilbert's projective metric): *Let \mathcal{C} be a proper cone with base \mathcal{B} . Then, for $b_1, b_2 \in \mathcal{B}$,*

$$\frac{1}{2} \|b_1 - b_2\|_{\mathcal{B}} = \mathcal{N}_{\mathcal{B}}(b_1 - b_2) \leq \tanh \frac{\mathfrak{h}_{\mathcal{C}}(b_1, b_2)}{4}. \quad (21)$$

More generally, if $\mathcal{B} = \mathcal{C} \cap H_e$, then for any $c_1, c_2 \in \mathcal{C}$ with $\langle e, c_2 \rangle \leq \langle e, c_1 \rangle$,

$$\mathcal{N}_{\mathcal{B}}(c_1 - c_2) \leq \langle e, c_2 \rangle \tanh \frac{\mathfrak{h}_{\mathcal{C}}(c_1, c_2)}{4}. \quad (22)$$

Proof: If $c_1 = 0$ or $c_2 = 0$ then $\langle e, c_2 \rangle = 0$, so that $\mathcal{N}_{\mathcal{B}}(c_1 - c_2) = 0$ and (22) holds. In all other cases, write $c_i =: \lambda_i b_i$ with $b_i \in \mathcal{B}$ and $\lambda_i := \langle e, c_i \rangle > 0$ for $i = 1, 2$ (for the proof of (21), set $\lambda_i := 1$ from the beginning). Define $m := \inf(b_1/b_2)$ and $M := \sup(b_1/b_2)$. If $m = 0$ or $M = \infty$ then $\mathfrak{h}_{\mathcal{C}}(c_1, c_2) = \mathfrak{h}(\lambda_1 b_1, \lambda_2 b_2) = \mathfrak{h}(b_1, b_2) = \infty$, and the statement follows from definition (18). Otherwise

$$mb_2 \leq_C b_1 \leq_C Mb_2,$$

which implies $b_1 - mb_2 \in \mathcal{C}$, so that $0 \leq \langle e, b_1 - mb_2 \rangle = 1 - m$; a similar reasoning for M gives $0 < m \leq 1 \leq M < \infty$. Now set,

$$F := \lambda_2 \frac{1-m}{M-m} b_1 + \lambda_2 \frac{Mm-m}{M-m} b_2, \quad (23)$$

(note that, if $m = M$, the statement (22) holds trivially since then $b_1 = b_2$), and write $c_1 - c_2 = (\lambda_1 b_1 - F) - (\lambda_2 b_2 - F)$. Observe that both expressions in parentheses are elements of \mathcal{C} , since

$$\lambda_1 b_1 - F \geq_C \lambda_2 b_1 - F = \lambda_2 \frac{M-1}{M-m} [b_1 - mb_2] \geq_C 0, \quad (24)$$

where the first inequality uses $\lambda_1 = \langle e, c_1 \rangle \geq \langle e, c_2 \rangle = \lambda_2$, and

$$\lambda_2 b_2 - F = \lambda_2 \frac{1-m}{M-m} [Mb_2 - b_1] \geq_C 0.$$

Thus, the difference representation $c_1 - c_2 = (\lambda_1 b_1 - F) - (\lambda_2 b_2 - F)$ occurs in the infimum in definition (18), and therefore, using $\langle e, b_1 \rangle = \langle e, b_2 \rangle = 1$,

$$\begin{aligned} \mathcal{N}_{\mathcal{B}}(c_1 - c_2) &\leq \langle e, \lambda_2 b_2 - F \rangle \\ &= \lambda_2 \frac{M+m-(1+Mm)}{M-m} \end{aligned} \quad (25)$$

$$\leq \lambda_2 \frac{M+m-2\sqrt{Mm}}{M-m} \quad (26)$$

$$\begin{aligned} &= \lambda_2 \frac{\sqrt{M}-\sqrt{m}}{\sqrt{M}+\sqrt{m}} \\ &= \langle e, c_2 \rangle \tanh [\mathfrak{h}_{\mathcal{C}}(c_1, c_2)/4], \end{aligned} \quad (27)$$

with Hilbert's projective metric $\mathfrak{h}_{\mathcal{C}}(c_1, c_2) = \mathfrak{h}(\lambda_1 b_1, \lambda_2 b_2) = \mathfrak{h}(b_1, b_2) = \ln(M/m)$. \square

Remark: Bounds stronger than in Proposition 7 hold when expressed directly in terms of the $\sup_{\mathcal{C}}$ and $\inf_{\mathcal{C}}$ used to define $\mathfrak{h}_{\mathcal{C}}$. For example, starting from (25) and continuing with elementary inequalities, for all $b_1, b_2 \in \mathcal{B}$,

$$\frac{1}{2} \|b_1 - b_2\|_{\mathcal{B}} = \mathcal{N}_{\mathcal{B}}(b_1 - b_2) \leq \frac{(\sup(b_1/b_2) - 1)(1 - \inf(b_1/b_2))}{\sup(b_1/b_2) - \inf(b_1/b_2)} \quad (28)$$

$$\leq \frac{1}{1 + \inf(b_1/b_2)} - \frac{1}{1 + \sup(b_1/b_2)} \quad (29)$$

$$\leq \tanh \frac{\mathfrak{h}_{\mathcal{C}}(b_1, b_2)}{4} \quad (30)$$

(despite appearance, all of these expressions are symmetric in b_1 and b_2).

Note further that b_1, b_2 and c_1, c_2 from Proposition 7 need to be elements of the cone \mathcal{C} in order for Hilbert's projective metric in (21) and (22) to be defined, whereas the lhs. of these inequalities depends only on the *differences* $b_1 - b_2$ and $c_1 - c_2$, respectively.

IV. CONTRACTIVITY PROPERTIES OF POSITIVE MAPS

We now relate the Hilbert metric contractivity properties of positive maps, in particular their projective diameter (11), to the contraction of base norms and negativities under application of the map. See also Theorem 4 (Birkhoff-Hopf contraction theorem), which is in the same spirit as the following.

In quantum information theory, given a quantum channel T and density matrices ρ_1, ρ_2 , the well-known contraction of the trace distance³³ implies that two quantum states do not become more distinguishable under the action of a channel,

$$\|T(\rho_1) - T(\rho_2)\|_1 \leq \|\rho_1 - \rho_2\|_1. \quad (31)$$

In the following, we will show that the rhs. of inequality (31) can be multiplied with a contraction factor $\eta \in [0, 1]$ that depends on the projective diameter $\Delta(T)$ of T . And we will generalize this to other base norms, some of which correspond to entanglement measures in quantum information theory and satisfy an analogue of (31) for LOCC channels T .^{36,37}

The setup will be that of linear maps $T : \mathcal{V} \rightarrow \mathcal{V}'$ between finite-dimensional vector spaces that contain proper cones $\mathcal{C} \subset \mathcal{V}$ and $\mathcal{C}' \subset \mathcal{V}'$, equipped, where necessary, with bases $\mathcal{B} = \mathcal{C} \cap H_e$ and $\mathcal{B}' = \mathcal{C}' \cap H_{e'}$, respectively. Recalling from Sec. II, T is called *cone-preserving*, or *positive*, if it preserves the property of an element lying in the cone, i.e., $T(\mathcal{C}) \subseteq \mathcal{C}'$. For several theorems, the stronger requirement for T to be *base-preserving* will be needed, meaning $T(\mathcal{B}) \subseteq \mathcal{B}'$. As \mathcal{B} spans the whole vector space \mathcal{V} , T is base-preserving if and only if T is cone-preserving and satisfies $\langle e, v \rangle = \langle e', T(v) \rangle$ for all $v \in \mathcal{V}$.

If T is linear and base-preserving, we immediately get that the base norm and the negativity contract under the application of T . That is, for all $v \in \mathcal{V}$,

$$\|T(v)\|_{\mathcal{B}'} \leq \|v\|_{\mathcal{B}}, \quad (32)$$

$$\mathcal{N}_{\mathcal{B}'}(T(v)) \leq \mathcal{N}_{\mathcal{B}}(v), \quad (33)$$

because whenever a representation $v = c_+ - c_-$ with $c_{\pm} \in \mathcal{C}$ occurs in the infimum (16) defining $\|v\|_{\mathcal{B}}$, then $T(v) = T(c_+) - T(c_-)$ is a valid representation for $T(v)$ as $T(c_{\pm}) \in \mathcal{C}'$ and one has $\langle e', T(c_{\pm}) \rangle = \langle e, c_{\pm} \rangle$; similarly for (33). The main results in this section will put contraction factors into (33) and (32) which depend on the projective diameter $\Delta(T)$ of the map T .

Paradigmatic application. Linear maps that are positive, in particular preserving the cone of positive semidefinite matrices \mathcal{S}_+ , are ubiquitous in quantum information theory. In this context one often considers, more restrictively, *completely positive* maps.²⁸ Many results, however, also hold for merely positive maps, or, more generally, for maps preserving other cones like $\mathcal{C} = \mathcal{S}_{\text{SEP}}, \mathcal{S}_{\text{PPT}}, \mathcal{S}_{\text{PPT}^+}$ (cf., example in Sec. III). Any physically realizable action on a quantum system corresponds to a map T that preserves the cone \mathcal{S}_+ , whereas more restricted actions preserve other cones as well. For instance, local quantum operations on a bipartite system with the possibility of classical communication between both sides (LOCC operations) preserve all of the cones mentioned above.

The requirement for a linear map T in quantum information theory to be *trace-preserving* (i.e., $\text{tr}[T(\rho)] = \text{tr}[\rho]$ for all density matrices $\rho \in \mathcal{B}_+$) translates to the requirement that $\langle e, v \rangle = \langle e', T(v) \rangle$ for all $v \in \mathcal{V}$, where again $e, e' = 1$ correspond to the usual trace on the respective spaces. This property is therefore weaker than the base-preserving property, which is equivalent to being positive *and* trace-preserving. However, for modeling a quantum operation that can either

succeed or fail, one usually employs a map T that is positive but not necessarily trace-preserving, interpreting $\text{tr}[T(\rho)]$ as the probability of success upon input of the state ρ ;²⁸ cf., Proposition 13.

We are now in a position to relate the contraction of the negativity and of the base norm under a map T to its projective diameter $\Delta(T)$.

Proposition 8 (Negativity contraction): Let $T : \mathcal{V} \rightarrow \mathcal{V}'$ be linear and base-preserving with respect to bases $\mathcal{B} = \mathcal{C} \cap H_e$ and \mathcal{B}' of proper cones $\mathcal{C} \subset \mathcal{V}$ and $\mathcal{C}' \subset \mathcal{V}'$, and let $v \in \mathcal{V}$ with $\langle e, v \rangle \geq 0$. Then

$$\mathcal{N}_{\mathcal{B}'}(T(v)) \leq \mathcal{N}_{\mathcal{B}}(v) \tanh \frac{\Delta(T)}{4}. \quad (34)$$

Proof: The proof is very similar to that of Proposition 7. According to (18), let $v = \lambda_1 b_1 - \lambda_2 b_2$ with $\lambda_2 = \mathcal{N}_{\mathcal{B}}(v)$, $b_1, b_2 \in \mathcal{B}$, and note $\lambda_1 \geq \lambda_2$ due to $\langle e, v \rangle \geq 0$. For $\Delta(T) = \infty$ the statement follows from (33), otherwise define $m := \inf(T(b_1)/T(b_2))$ and $M := \sup(T(b_1)/T(b_2))$. Again $0 < m \leq 1 \leq M < \infty$, since $T(b_1), T(b_2) \in \mathcal{B}'$ and

$$mT(b_2) \leq_{\mathcal{C}'} T(b_1) \leq_{\mathcal{C}'} MT(b_2). \quad (35)$$

Defining

$$F := \lambda_2 \frac{1-m}{M-m} T(b_1) + \lambda_2 \frac{Mm-m}{M-m} T(b_2),$$

writing $T(v) = (\lambda_1 T(b_1) - F) - (\lambda_2 T(b_2) - F)$ and repeating the steps from (24) to (27) yields

$$\begin{aligned} \mathcal{N}_{\mathcal{B}'}(T(v)) &\leq \mathcal{N}_{\mathcal{B}}(v) \tanh [\mathfrak{h}_{\mathcal{C}'}(T(b_1), T(b_2))/4] \\ &\leq \mathcal{N}_{\mathcal{B}}(v) \tanh [\Delta(T)/4]. \end{aligned} \quad (36)$$

□

Corollary 9 (Contraction of base norm distance): Let $T : \mathcal{V} \rightarrow \mathcal{V}'$ be linear and base-preserving with respect to bases $\mathcal{B} = \mathcal{C} \cap H_e$ and \mathcal{B}' of proper cones $\mathcal{C} \subset \mathcal{V}$ and $\mathcal{C}' \subset \mathcal{V}'$, and let $v_1, v_2 \in \mathcal{V}$ with $\langle e, v_1 \rangle = \langle e, v_2 \rangle$. Then

$$\|T(v_1) - T(v_2)\|_{\mathcal{B}'} \leq \|v_1 - v_2\|_{\mathcal{B}} \tanh \frac{\Delta(T)}{4}. \quad (37)$$

Proof: Note $\|v_1 - v_2\|_{\mathcal{B}} = 2\mathcal{N}_{\mathcal{B}}(v_1 - v_2)$ and $\|T(v_1 - v_2)\|_{\mathcal{B}'} = 2\mathcal{N}_{\mathcal{B}'}(T(v_1 - v_2))$ since $\langle e, v_1 - v_2 \rangle = \langle e', T(v_1 - v_2) \rangle = 0$, and use Proposition 8. □

In the context of quantum information theory, we get a potentially non-trivial contraction of the trace-norm when applied to a *difference* of quantum states, which is the usual situation in state discrimination,

$$\|T(\rho_1) - T(\rho_2)\|_1 \leq \|\rho_1 - \rho_2\|_1 \tanh \frac{\Delta(T)}{4}. \quad (38)$$

If $\Delta(T) < \infty$, this improves Ruskai's trace-norm contraction inequality (31). $\Delta(T)$ is finite in particular if the image $T(\mathcal{C})$ lies in the interior of the cone \mathcal{C}' , for instance, if T maps every state to a full-rank density matrix. We will expand further on distinguishability measures in Sec. V.

However, under the general conditions of Proposition 8, a non-trivial contraction result for the base norm cannot exist, since for base-preserving T we have $\|T(v)\|_{\mathcal{B}'} = \|v\|_{\mathcal{B}} > 0$ for all $v \in \mathcal{C} \setminus 0$,

i.e., there cannot be strict contraction. This explains the necessity for an additional condition (like $T(v) \notin C'$) in the following proposition and also the different contraction coefficient:

Proposition 10 (Base norm contraction; Logarithmic negativity decrease): Let $T : \mathcal{V} \rightarrow \mathcal{V}'$ be linear and base-preserving with respect to bases $\mathcal{B} = \mathcal{C} \cap H_e$ and \mathcal{B}' of proper cones $\mathcal{C} \subset \mathcal{V}$ and $\mathcal{C}' \subset \mathcal{V}'$, and let $v \in \mathcal{V}$ with $\langle e, v \rangle \geq 0$. If $T(v) \notin \mathcal{C}'$, then

$$\|T(v)\|_{\mathcal{B}'} \leq \|v\|_{\mathcal{B}} \tanh \frac{\Delta(T)}{2}. \quad (39)$$

Proof: The idea is the same as in the proof of Proposition 8, but now, in the same notation, use for subtraction the linear combination,

$$F := \frac{\lambda_2}{M + 1/m} T(b_1) + \frac{\lambda_1}{M + 1/m} T(b_2).$$

If $\lambda_2/\lambda_1 \leq m$, then $T(v) = \lambda_1 [T(b_1) - (\lambda_2/\lambda_1)T(b_2)] \geq_{\mathcal{C}'} 0$ due to (35), i.e., $T(v) \in \mathcal{C}'$ contrary to assumption; the same contradiction is obtained for $\lambda_1 = 0$, as this would imply $v = 0$. Therefore $\lambda_2 > m\lambda_1$, which ensures that both terms in the difference representation $T(v) = (\lambda_1 T(b_1) - F) - (\lambda_2 T(b_2) - F)$ are non-negative,

$$\begin{aligned} \lambda_1 T(b_1) - F &= \frac{1}{M + 1/m} \left[(M\lambda_1 - \lambda_2)T(b_1) + \frac{\lambda_1}{m}(T(b_1) - mT(b_2)) \right] \geq_{\mathcal{C}'} 0, \\ \lambda_2 T(b_2) - F &= \frac{1}{M + 1/m} \left[\frac{1}{m}(\lambda_2 - m\lambda_1)T(b_2) + \lambda_2(MT(b_2) - T(b_1)) \right] \geq_{\mathcal{C}'} 0. \end{aligned}$$

Thus, from definition (16),

$$\begin{aligned} \|T(v)\|_{\mathcal{B}'} &\leq \langle e', \lambda_1 T(b_1) - F \rangle + \langle e', \lambda_2 T(b_2) - F \rangle \\ &= (\lambda_1 + \lambda_2) \left(1 - \frac{2}{M + 1/m} \right) \\ &\leq \|v\|_{\mathcal{B}} \frac{(M + 1/m - 2)(M + m) + 2(M - 1)(1 - m)}{(M + 1/m)(M + m)} \\ &= \|v\|_{\mathcal{B}} \frac{M - m}{M + m} = \|v\|_{\mathcal{B}} \tanh[\mathfrak{h}_{\mathcal{C}'}(T(b_1), T(b_2))/2] \\ &\leq \|v\|_{\mathcal{B}} \tanh[\Delta(T)/2], \end{aligned}$$

where the third line becomes an inequality since the non-negative term $2(M - 1)(1 - m)$ was added to the numerator, and we used $\|v\|_{\mathcal{B}} = \lambda_1 + \lambda_2$ due to the choice of λ_2 . \square

One obvious consequence of (39) is an *additive* decrease of the logarithmic negativity $\log \|v\|_{\mathcal{B}}$, which is the quantity that more naturally appears in entanglement theory. Another implication is the following:

Corollary 11 (Contraction into cone in finite time): Using the same proper cone $\mathcal{C} \subset \mathcal{V}$ and base $\mathcal{B} = \mathcal{C} \cap H_e$ in both the domain and codomain, let $T : \mathcal{V} \rightarrow \mathcal{V}$ be a linear and base-preserving map. Let $v \in \mathcal{V}$ with $\langle e, v \rangle = 1$. Then, $T^n(v) \in \mathcal{C}$ for any $n \in \mathbb{N}$ with

$$n \geq \frac{\log \|v\|_{\mathcal{B}}}{-\log \tanh[\Delta(T)/2]}. \quad (40)$$

Another, albeit weaker, sufficient condition for $T^n(v) \in \mathcal{C}$ is

$$n \geq \frac{e^{\Delta(T)}}{2} \ln \|v\|_{\mathcal{B}}. \quad (41)$$

Proof: By contradiction: Let n satisfy (40) and assume $T^n(v) \notin \mathcal{C}$. Then, since T is cone-preserving, $T^k(v) \notin \mathcal{C}$ for all $k = 1, \dots, n$, and Proposition 10 can be applied n times:

$$\begin{aligned} \log(\|T^n(v)\|_{\mathcal{B}}) &\leq \log((\tanh[\Delta(T)/2])^n \|v\|_{\mathcal{B}}) \\ &= \log\|v\|_{\mathcal{B}} + n \log \tanh[\Delta(T)/2] \leq 0, \end{aligned}$$

i.e., $\|T^n(v)\|_{\mathcal{B}} \leq 1 = \langle e, v \rangle = \langle e, T^n(v) \rangle$. This implies $T^n(v) \in \mathcal{C}$, which is the desired contradiction.

(41) is a more restrictive condition on n than (40), since $-\ln \tanh[\Delta(T)/2] \geq 2/e^{\Delta(T)}$ which follows from

$$-\frac{2}{e^{\Delta(T)}} - \ln \tanh \frac{\Delta(T)}{2} = \int_{\Delta(T)}^{\infty} dx \frac{d}{dx} \left(\frac{2}{e^x} + \ln \tanh \frac{x}{2} \right) = \int_{\Delta(T)}^{\infty} dx \frac{2e^{-2x}}{e^x - e^{-x}} \geq 0. \quad \square$$

One might wonder whether the contraction factors in the previous propositions, Eqs. (34), (37), and (39) are optimal and why the hyperbolic tangent appears. In Appendix C, we show that the contraction factors are indeed the best possible, provided that they are to depend only on $\Delta(T)$ but not on other characteristics of T . Also, the upper bounds in Proposition 7 are tight if they are to depend only on the Hilbert distance.

The following proposition formalizes the *contraction ratio* $\eta^b(T)$ of a linear map T , not required to be base- or cone-preserving, with respect to base norms. This statement was noted before in Ref. 33 for the trace-norm $\|\cdot\|_1 = \|\cdot\|_{\mathcal{B}_+}$, in which case the extreme points, $\text{ext}(\mathcal{B}_+)$, are the *pure quantum states*.

Proposition 12 (Base norm contraction coefficient): Let $T : \mathcal{V} \rightarrow \mathcal{V}'$ be a linear map, and let \mathcal{B} and \mathcal{B}' be bases of proper cones $\mathcal{C} \subset \mathcal{V}$ and $\mathcal{C}' \subset \mathcal{V}'$. Then,

$$\eta^b(T) := \sup_{v_1 \neq v_2 \in \mathcal{B}} \frac{\|T(v_1) - T(v_2)\|_{\mathcal{B}'}}{\|v_1 - v_2\|_{\mathcal{B}}} = \frac{1}{2} \sup_{v_1, v_2 \in \text{ext}(\mathcal{B})} \|T(v_1) - T(v_2)\|_{\mathcal{B}'}. \quad (42)$$

The supremum on the right can be taken alternatively also over all points in the base, $v_1, v_2 \in \mathcal{B}$.

Proof: Choose any $v_1 \neq v_2 \in \mathcal{B}$ and let $v_1 - v_2 =: \lambda_1 b_1 - \lambda_2 b_2$ such that $\|v_1 - v_2\|_{\mathcal{B}} = \lambda_1 + \lambda_2$ and $b_1, b_2 \in \mathcal{B}$. Note that $0 < \lambda_1 = \lambda_2 =: \lambda \leq 1$ and therefore $\|v_1 - v_2\|_{\mathcal{B}} = 2\lambda \leq 2$, so that $\|T(v_1 - v_2)\|_{\mathcal{B}'} / \|v_1 - v_2\|_{\mathcal{B}} \geq \|T(v_1 - v_2)\|_{\mathcal{B}'}/2$ which shows that the rhs. in (42) is certainly a lower bound.

To prove inequality in the other direction, note that, in the same notation, $v_1 - v_2 = \lambda(b_1 - b_2)$ and $\|b_1 - b_2\|_{\mathcal{B}} = \|v_1 - v_2\|_{\mathcal{B}}/\lambda = 2$, and thus $\|T(v_1 - v_2)\|_{\mathcal{B}'} / \|v_1 - v_2\|_{\mathcal{B}} = \|T(b_1 - b_2)\|_{\mathcal{B}'}/2$. By Caratheodory's theorem, b_1 and b_2 can each be written as a convex combination of finitely many extreme points $b_1^{(i)}, b_2^{(i)} \in \text{ext}(\mathcal{B})$, so that in a common expansion with $\sum_i \mu_i = 1, \mu_i \geq 0$,

$$\frac{\|T(v_1) - T(v_2)\|_{\mathcal{B}'}}{\|v_1 - v_2\|_{\mathcal{B}}} = \frac{1}{2} \left\| \sum_i \mu_i T(b_1^{(i)} - b_2^{(i)}) \right\|_{\mathcal{B}'} \leq \frac{1}{2} \sum_i \mu_i \|T(b_1^{(i)} - b_2^{(i)})\|_{\mathcal{B}'}. \quad (43)$$

Thus, there exists an index i such that $\|T(b_1^{(i)} - b_2^{(i)})\|_{\mathcal{B}'}/2$ is greater than or equal to the lhs, proving (42). \square

Proposition 12 connects the two very similar proofs of Proposition 7 and Corollary 9, as it allows to prove the latter from the former,

$$\begin{aligned} \sup_{v_1, v_2} \frac{\|T(v_1) - T(v_2)\|_{\mathcal{B}'}}{\|v_1 - v_2\|_{\mathcal{B}}} &= \frac{1}{2} \sup_{b_1, b_2 \in \mathcal{B}} \|T(b_1) - T(b_2)\|_{\mathcal{B}'} \\ &\leq \sup_{b_1, b_2 \in \mathcal{B}} \tanh[\mathfrak{h}_{\mathcal{C}'}(T(b_1), T(b_2))/4] \\ &= \tanh[\Delta(T)/4], \end{aligned} \quad (43)$$

where the first equality is Proposition 12 (the first supremum runs over all pairs $v_1 \neq v_2 \in \mathcal{V}$ with $\langle e, v_1 \rangle = \langle e, v_2 \rangle$) and the inequality follows from Proposition 7.

By the Birkhoff-Hopf theorem (Theorem 4), the contraction ratios of Hilbert's projective metric, and the oscillation are $\eta^b(T) = \eta^{\text{osc}}(T) = \tanh[\Delta(T)/4]$. Corollary 9 or Eq. (43) shows that $\eta^b(T) \leq \eta^b(T)$ for base-preserving T . In Appendix A, for qubit channels and with respect to the positive semidefinite cone \mathcal{S}_+ , we obtain a characterization of the cases where the trace-norm contraction coefficient actually equals $\tanh[\Delta(T)/4]$ (Proposition 22).

From the defining equations (8) and (11), it is apparent that the diameter $\Delta_{\mathcal{C} \rightarrow \mathcal{C}'}(T)$ of $T : \mathcal{C} \rightarrow \mathcal{C}'$ decreases or stays constant when \mathcal{C} is being restricted to a subcone $\mathcal{D} \subseteq \mathcal{C}$, i.e., $\Delta_{\mathcal{D} \rightarrow \mathcal{C}'}(T) \leq \Delta_{\mathcal{C} \rightarrow \mathcal{C}'}(T)$, and that it increases or stays constant when \mathcal{C}' is being restricted to a subcone $\mathcal{D}' \subseteq \mathcal{C}'$, i.e., $\Delta_{\mathcal{C} \rightarrow \mathcal{D}'}(T) \geq \Delta_{\mathcal{C} \rightarrow \mathcal{C}'}(T)$. At the end of Appendix A we show by way of examples that there is no such monotonicity of the projective diameter in the common case where both cones $\mathcal{C} = \mathcal{C}'$ are identical and varied simultaneously.

Examples: In Appendix A we look at Hilbert's projective metric in the state space of a qubit, also for different choices of cones, and connect the projective diameter to the trace-norm contraction coefficient. In Appendix B we compute the projective diameter of some general depolarizing channels, also commenting on a bipartite scenario.

As mentioned earlier in this section, positive maps T that are not necessarily trace-preserving are used in quantum theory to model operations on a quantum system which do not succeed with certainty, but instead with some probability $p = \text{tr}[T(\rho)]$. In this context one often requires one operation out of a collection $\{T_i\}$ of possible operations to succeed with certainty, and one interprets $\rho_i := T_i(\rho)/p_i$ as the state of the system after the occurrence of operation i . A direct analogue of Proposition 8 does not hold for maps T_i that do not preserve normalization; in an averaged sense, however, contraction does still occur as we now show. As it is primarily inspired by the physical context, we will partly use quantum theoretical notation for the following proposition and discuss its meaning afterwards. The statement holds, however, for general cones and bases.

Proposition 13 (Negativity contraction under non-deterministic operations): Let $T_i : \mathcal{V} \rightarrow \mathcal{V}_i$ with $i = 1, \dots, N$ be linear and cone-preserving maps with respect to proper cones $\mathcal{C} \subset \mathcal{V}$ and $\mathcal{C}_i \subset \mathcal{V}_i$ with bases $\mathcal{B} = \mathcal{C} \cap H_e$ and $\mathcal{B}_i = \mathcal{C}_i \cap H_{e_i}$, satisfying $\sum_{i=1}^N \langle e_i, T_i(b) \rangle \leq 1$ for all $b \in \mathcal{B}$. Let $\rho \in \mathcal{V}$ with $p_i := \langle e_i, T_i(\rho) \rangle \geq 0$. Then,

$$\sum_{i=1}^N p_i \mathcal{N}_{\mathcal{B}_i}(\rho_i) \leq \mathcal{N}_{\mathcal{B}}(\rho) \tanh \frac{\max_i \Delta(T_i)}{4}, \quad (44)$$

for any $\rho_i \in \mathcal{V}_i$ that satisfy $T_i(\rho) = p_i \rho_i$ whenever $p_i > 0$.

Proof: Similar to the proofs of Propositions 7 and 8, let $\rho = \lambda_1 b_1 - \lambda_2 b_2$ with $\lambda_2 = \mathcal{N}_{\mathcal{B}}(\rho)$ and $b_1, b_2 \in \mathcal{B}$. Thus,

$$T_i(\rho) = \lambda_1 \langle e_i, T_i(b_1) \rangle b_1^{(i)} - \lambda_2 \langle e_i, T_i(b_2) \rangle b_2^{(i)}$$

for some $b_1^{(i)}, b_2^{(i)} \in \mathcal{B}_i$. Setting $m_i := \inf(b_1^{(i)}/b_2^{(i)})$, $M_i := \sup(b_1^{(i)}/b_2^{(i)})$ and

$$F_i := \lambda_2 \langle e_i, T_i(b_2) \rangle \left[(1 - m_i) b_1^{(i)} + m_i (M_i - 1) b_2^{(i)} \right] / (M_i - m_i),$$

and using $0 \leq p_i = \lambda_1 \langle e_i, T_i(b_1) \rangle - \lambda_2 \langle e_i, T_i(b_2) \rangle$ and $\lambda_2 = \mathcal{N}_{\mathcal{B}}(\rho)$, one arrives at the equivalent of (36):

$$\mathcal{N}_{\mathcal{B}_i}(T_i(\rho)) \leq \mathcal{N}_{\mathcal{B}}(\rho) \langle e_i, T_i(b_2) \rangle \tanh [\Delta(T_i)/4]$$

for each $i = 1, \dots, N$. By disregarding terms with $p_i = 0$, this yields

$$\begin{aligned} \sum_i p_i \mathcal{N}_{\mathcal{B}_i}(\rho_i) &= \sum_i \mathcal{N}_{\mathcal{B}_i}(p_i \rho_i) = \sum_i \mathcal{N}_{\mathcal{B}_i}(T_i(\rho)) \\ &\leq \mathcal{N}_{\mathcal{B}}(\rho) \sum_i \langle e_i, T_i(b_2) \rangle \tanh [\Delta(T_i)/4] \end{aligned}$$

$$\leq \mathcal{N}_B(\rho) \max_j \tanh[\Delta(T_j)/4] \sum_i \langle e_i, T_i(b_2) \rangle,$$

and the last sum is at most 1 by assumption. \square

In entanglement theory, replacing the hyperbolic tangent in (44) by 1 gives exactly the requirement for \mathcal{N}_B to be an *entanglement monotone*, when considered for normalized quantum states ρ .¹⁹ As $\sum_i p_i = 1$ and $p_i \geq 0$ in this case, the general base norm and the logarithmic negativity are entanglement monotones as well,^{30,37}

$$\sum_i p_i \|\rho_i\|_{B_i} = \sum_i p_i (2\mathcal{N}_{B_i}(\rho_i) + 1) \leq 2\mathcal{N}_B(\rho) + \sum_i p_i = \|\rho\|_B, \quad (45)$$

$$\sum_i p_i \log \|\rho_i\|_{B_i} \leq \log \left(\sum_i p_i \|\rho_i\|_{B_i} \right) \leq \log \|\rho\|_B. \quad (46)$$

Proposition 13 yields a (potentially) non-trivial contraction ratio in the inequality that shows the (generalized) negativity to be an entanglement monotone.³⁷ But putting non-trivial contraction coefficients that depend solely on the projective diameter into Eqs. (45) and (46) would require some additional assumptions akin to Proposition 10, such as $T_i(\rho) \notin \mathcal{C}_i$ for all i , which, however, seems very restrictive in the context here. But with this additional requirement, the lhs. of (46), for instance, is upper bounded by $\log \|\rho\|_B + \log \tanh[\max_i \Delta(T_i)/2]$.

Note that, similar to the trace-norm, the base norm associated with the cone S_{PPT} also has a physical interpretation. For bipartite quantum systems, the logarithmic negativity $\log \|\rho\|_{S_{\text{PPT}}}$ is an upper bound to the distillable entanglement,³⁷ while it is a lower bound to the PPT-entanglement cost and, for many states ρ , it exactly equals the latter.³ Proposition 10 therefore states that under the application of a PPT-channel T the upper bound on the distillable entanglement of a quantum state ρ will decrease by at least $\log \tanh[\Delta(T)/2]$ unless $T(\rho)$ is not distillable in the first place; note that, for the normalized quantum state $T(\rho)$, the condition $T(\rho) \in S_{\text{PPT}}$ is equivalent to $\log \|T(\rho)\|_{S_{\text{PPT}}} = 0$. And for repeated applications of the PPT-channel T , Corollary 11 implies that the state $T^n(\rho)$ after $n \geq (e^{\Delta(T)/2}) \ln \|\rho^{T_1}\|_1$ time steps will not be distillable at all and that its PPT entanglement cost will vanish.

V. DISTINGUISHABILITY MEASURES

In the preceding sections we have established relations between Hilbert's projective metric and base norms and negativities, tools that are used in quantum information theory to quantify entanglement in a bipartite quantum system. And apart from representing merely abstract measures quantifying the distance between a quantum state and a given cone, they also give upper bounds on physical quantities, like the distillable entanglement, for some special choices of cones.^{3,37}

Another physical interpretation, which was already insinuated above in (31) and (38), is that the trace distance $\|\rho_1 - \rho_2\|_1 = \|\rho_1 - \rho_2\|_{B_+}$ quantifies the best possible distinguishability between two quantum states ρ_1, ρ_2 under all physical measurements.^{16,21,28} In this section we will make this notion precise by developing a similar duality relation between more general distinguishability measures on the one hand and base norms associated with general cones on the other hand, and relate these to Hilbert's projective metric. In several of the results from Secs. III and IV, the base norm is naturally applied to a *difference* of two elements, and it is these results that are most readily translated to distinguishability measures, which we will do later in this section.

The following setting is inspired by physical considerations and will be explicitly translated into the quantum information context below Theorem 14 (for more on those distinguishability measures, see also Ref. 27). Let \mathcal{V} be a finite-dimensional real vector space and \mathcal{V}^* its dual, equipped with a distinguished element $e \in \mathcal{V}^*$. Furthermore, let $M \subset \mathcal{V}^*$ be a closed convex set with non-empty interior which satisfies

$$M \cap (-M) = \{0\} \quad \text{and} \quad e - M \subseteq M \quad (47)$$

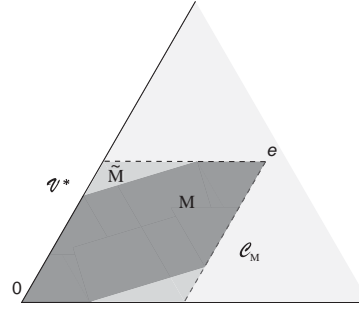


FIG. 1. A 2-dimensional section, containing the origin 0 and the distinguished element $e \in \mathcal{V}^*$, through the cone \mathcal{C}_M . The set M (dark shade) can in general be a proper subset of \tilde{M} (including the lighter shade), which contains the elements $E \in \mathcal{V}^*$ that satisfy $0 \leq E \leq e$, see (50). Both M and \tilde{M} generate the same cone $\mathcal{C}_M = \mathcal{C}_{\tilde{M}}$ (including the lightest shade).

(where the latter means: $E \in M \Rightarrow e - E \in M$); see Fig. 1 for illustration. A set M satisfying these conditions generates a proper cone, which we will write as $\mathcal{C}_M := \bigcup_{\lambda \geq 0} \lambda M$. Denoting its dual cone by $\mathcal{C} := (\mathcal{C}_M)^* \subset \mathcal{V}$, \mathcal{C}_M and \mathcal{C} induce partial orders $\leq_{\mathcal{C}_M}$ and $\leq_{\mathcal{C}}$ in \mathcal{V}^* and \mathcal{V} , respectively. Define for $v \in \mathcal{V}$:

$$[v]_{(M)} := \sup_{E \in M} \langle E, v \rangle, \quad (48)$$

$$\|v\|_{(M)} := 2[v]_{(M)} - \langle e, v \rangle = \sup_{E \in M} \langle 2E - e, v \rangle. \quad (49)$$

Under the above conditions, the last line defines a norm $\|\cdot\|_{(M)}$ on \mathcal{V} ,²⁷ here called the *distinguishability norm* (notice the difference in notation between (49) and the base norm (15)). Note, however, that $[\cdot]_{(M)}$ in (48) does *not* define a norm on \mathcal{V} , as, for instance, $[v]_{(M)} = 0$ for all $v \in (-\mathcal{C})$. Furthermore, starting from the set M above, define (cf., Fig. 1)

$$\tilde{M} := \{E \in \mathcal{V}^* \mid 0 \leq_{\mathcal{C}_M} E \leq_{\mathcal{C}_M} e\}. \quad (50)$$

The following is the main theorem in this section and establishes first that also $\|\cdot\|_{(\tilde{M})}$ is a well-defined distinguishability norm, whose distinguishing power, in the context of quantum information theory, is at least as good as that of $\|\cdot\|_{(M)}$. The theorem then relates both of these distinguishability norms to a base norm $\|\cdot\|_{\mathcal{B}}$ on \mathcal{V} , where the base \mathcal{B} is defined by the cone \mathcal{C} and the functional e from above.

Theorem 14 (Duality between distinguishability norms and base norms): *Under the above conditions, \tilde{M} from (50) contains M , it generates the same cone as M , i.e., $\mathcal{C}_M = \mathcal{C}_{\tilde{M}} := \bigcup_{\lambda \geq 0} \lambda \tilde{M}$, and it induces a well-defined distinguishability norm $\|\cdot\|_{(\tilde{M})}$ via (49). Furthermore, $\mathcal{B} := \mathcal{C} \cap H_e$ is a base of the cone $\mathcal{C} := (\mathcal{C}_M)^*$ and therefore induces a well-defined base norm $\|\cdot\|_{\mathcal{B}}$ on \mathcal{V} via (15). These distinguishability and base norms satisfy*

$$\|v\|_{(M)} \leq \|v\|_{(\tilde{M})} = \|v\|_{\mathcal{B}} \quad (51)$$

and

$$\sup_{E \in M} \langle E, v \rangle \leq \sup_{0 \leq E \leq e} \langle E, v \rangle = \frac{1}{2} (\|v\|_{\mathcal{B}} + \langle e, v \rangle) \quad (52)$$

for all $v \in \mathcal{V}$.

Proof: $E \in M$ implies $E \in \mathcal{C}_M$ and $e - E \in M \subset \mathcal{C}_M$, so that $0 \leq E \leq e$ and $E \in \tilde{M}$ according to (50), which shows $M \subseteq \tilde{M}$ and subsequently $\mathcal{C}_M \subseteq \mathcal{C}_{\tilde{M}}$. On the other hand, $E \in \mathcal{C}_{\tilde{M}}$ means $E = \lambda E'$ for some $\lambda \geq 0$ and $E' \in \tilde{M}$, so in particular $E' \in \mathcal{C}_M$; this implies $E = \lambda E' \in \lambda \mathcal{C}_M \subseteq \mathcal{C}_M$, so that also $\mathcal{C}_{\tilde{M}} \subseteq \mathcal{C}_M$. Since the cone \mathcal{C}_M appearing in definition (50) is proper, \tilde{M} is closed and convex and satisfies $e - \tilde{M} \subseteq \tilde{M}$; also, $E \in \tilde{M} \cap (-\tilde{M})$ implies $E \leq 0 \leq E$, i.e., $E = 0$; lastly, due

to $\tilde{M} \supseteq M$, \tilde{M} contains 0 and has non-empty interior. Therefore, \tilde{M} has all the properties necessary to define a distinguishability norm via (49).

Next we will show that $\mathcal{B} = \mathcal{C} \cap \{v \in \mathcal{V} | \langle e, v \rangle = 1\}$ forms a base of \mathcal{C} . Note that, as an intersection of convex sets, \mathcal{B} is convex. Now we will show that $e \in (\mathcal{C}^*)^\circ$. As M has non-empty interior, it contains an open ball $U_\epsilon(a)$ of radius $\epsilon > 0$ around $a \in M$, i.e., $U_\epsilon(a) \subseteq M$. (47) then implies $e - U_\epsilon(a) = U_\epsilon(e - a) \subseteq M$, and convexity of M gives $\frac{1}{2}U_\epsilon(a) + \frac{1}{2}U_\epsilon(e - a) = U_\epsilon(e/2) \subseteq M$. Thus, $U_{2\epsilon}(e) = 2U_\epsilon(e/2) \subseteq 2M \subseteq \mathcal{C}_M$, so that $e \in (\mathcal{C}_M)^\circ = (\mathcal{C}^*)^\circ$. Now let $v \in \mathcal{C} \setminus 0$; we need to show that v can be written in a unique way as $v = \lambda b$ with $\lambda > 0$ and $b \in \mathcal{B}$. First, $e \in \mathcal{C}^*$ gives $\langle e, v \rangle \geq 0$. Now assume $\langle e, v \rangle = 0$. The function $\langle f, v \rangle$ is linear in $f \in \mathcal{V}^*$ and non-constant since $v \neq 0$, and therefore $\langle f, v \rangle < 0$ for some $f \in U_{2\epsilon}(e) \subseteq \mathcal{C}^*$, a contradiction. Thus $\langle e, v \rangle > 0$, and so $\lambda := \langle e, v \rangle$ and $b := v/\lambda$ give the desired unique representation $v = \lambda b$.

The inequality in (51) follows from $M \subseteq \tilde{M}$, and the equality follows from the strong duality between two semidefinite programs,¹⁰ each corresponding to one side of the equation. First, weak duality gives

$$\begin{aligned} \|v\|_{(\tilde{M})} &= \sup \{ \langle 2E, v \rangle - \langle e, v \rangle \mid E \geq_{\mathcal{C}_M} 0, e - E \geq_{\mathcal{C}_M} 0 \} \\ &\leq \inf \{ \langle e, 2c_+ \rangle - \langle e, v \rangle \mid v = c_+ - c_-, c_\pm \geq_{(\mathcal{C}_M)^*} 0 \} \\ &= \inf \{ \langle e, c_+ \rangle + \langle e, c_- \rangle \mid v = c_+ - c_-, c_\pm \in \mathcal{C} \} \\ &= \|v\|_{\mathcal{B}}. \end{aligned} \quad (53)$$

Since both \mathcal{C}_M and \mathcal{C} have non-empty interior, Slater's constraint qualification¹⁰ yields actually equality in (53) and ensures that all optima are attained. (52) follows from (51) and the definitions (49) and (50). \square

The construction above can also be reversed, albeit in a partially non-unique manner: Starting from a vector space \mathcal{V} with base norm $\|\cdot\|_{\mathcal{B}}$, where $\mathcal{B} = \mathcal{C} \cap H_e$ is a base of a proper cone $\mathcal{C} \subset \mathcal{V}$, one can identify $\mathcal{C}_M := \mathcal{C}^* \subset \mathcal{V}^*$ and then $e \in (\mathcal{C}_M)^\circ$ will hold. There are, however, different possible choices for M that all satisfy the conditions above, see Fig. 1. But each of these choices will lead to the same \tilde{M} , as by (50) this only depends on \mathcal{C}_M (and \tilde{M} itself is a possible choice for M). Theorem 14 and in particular the relations (51) and (52) also hold in this situation, and it is indeed the distinguishability norm associated with the unique \tilde{M} which is strongly dual to the base norm $\|\cdot\|_{\mathcal{B}}$, i.e., which attains equality in (51).

In the context of quantum theory, the space \mathcal{V}^* from above is the vector space of all Hermitian observables in $\mathcal{M}_d(\mathbb{C})$, including as the distinguished element the identity matrix $1 =: e$ which corresponds to the trace functional and acts on $A \in \mathcal{V}$ (the vector space containing the quantum states) by $\langle e, A \rangle = \text{tr}[1A] = \text{tr}[A]$, since \mathcal{V} is identified with the set of Hermitian matrices in $\mathcal{M}_d(\mathbb{C})$ by the Hilbert-Schmidt inner product. M is the set of POVM elements of 2-outcome POVMs that are realizable in a given physical setup.²⁸ For any $E \in M$, the probability of outcome E in the measurement corresponding to the POVM $(E, 1 - E)$ on a valid quantum state ρ is $\text{tr}[E\rho] = \langle E, \rho \rangle$. $E \in M$ is called an *effect operator* or *measurement operator*.

The above requirements on M derive directly from physical considerations (see also Theorem 4 in Ref. 27): (i) a convex combination of allowed measurements corresponds to their probabilistic mixture and is therefore also allowed, (ii) exactly when M has non-empty interior is it possible to reconstruct the quantum state ρ from the knowledge of all probabilities $\text{tr}[E\rho]$,²⁷ (iii) for each $E \in M$, by relabeling the two outcomes of the corresponding POVM $(E, 1 - E)$, also $(1 - E, E)$ is an implementable POVM, i.e., $1 - E \in M$, and (iv) the POVM $(0, 1)$ which yields the second outcome with probability 1 is trivially implementable, so $0 \in M$. As probabilities have to be non-negative, valid quantum states satisfy $\rho \in (\mathcal{C}_M)^* = \mathcal{C}$, and since the normalization of states is measured by the observable 1, all physical quantum states ρ are, in the present setting, necessarily elements of the base $\mathcal{B} = \mathcal{C} \cap H_1$. Note further, (v) that demanding non-negative probabilities for all states in a

set with non-empty interior requires $M \cap (-M) \subseteq \{0\}$. For quantum states ρ we also automatically have $\text{tr}[E\rho] \leq 1$ for all $E \in M$ since $1 - \text{tr}[E\rho] = \text{tr}[(1 - E)\rho] \geq 0$ due to $1 - E \in M$.

A basic task in quantum information theory is that of distinguishing two (*a priori* equiprobable) quantum states ρ_1, ρ_2 , i.e., finding the 2-outcome POVM $(E, 1 - E)$ in a set of implementable POVMs (corresponding to the set M) which maximizes the difference (*bias*) between the probabilities of outcome E when measuring on state ρ_1 versus ρ_2 . This *maximal bias*²⁷ is

$$\sup_{E \in M} (\text{tr}[E\rho_1] - \text{tr}[E\rho_2]) = \sup_{E \in M} \text{tr}[E(\rho_1 - \rho_2)] = [\rho_1 - \rho_2]_{(M)} = \frac{1}{2} \|\rho_1 - \rho_2\|_{(M)},$$

where the last equality holds due to $\text{tr}[\rho_1] = \text{tr}[\rho_2]$, cf. (48) and (49). Theorem 14 then gives the relation between these distinguishability measures and the base norm $\|\cdot\|_{\mathcal{C}}$:

$$\sup_{E \in M} \text{tr}[E(\rho_1 - \rho_2)] = \frac{1}{2} \|\rho_1 - \rho_2\|_{(M)} \leq \frac{1}{2} \|\rho_1 - \rho_2\|_{\mathcal{C}} = \frac{1}{2} \|\rho_1 - \rho_2\|_{(\mathcal{C}_M)^*}, \quad (54)$$

where the first and the last expression explicitly show the duality going from M to $\mathcal{C} = (\mathcal{C}_M)^*$. By Theorem 14, we have equality in (54) if $M = \tilde{M}$, which translates to the following condition in the quantum context (cf., Fig. 1): if $(E_i, 1 - E_i)$ are two implementable POVMs (i.e., if $E_i \in M$ for $i = 1, 2$) and if $\alpha_i > 1$ are numbers such that $\alpha_1 E_1 + \alpha_2 E_2 = 1$, then also the POVM $(\alpha_1 E_1, \alpha_2 E_2)$ is implementable (i.e., $\alpha_1 E_1 \in M$). Equality in (54) indeed holds for some important classes of measurements considered in quantum information theory, as we discuss now.

The best known instance, the set of POVM elements,

$$M_+ := \{E \in \mathcal{M}_d(\mathbb{C}) \mid E \in \mathcal{S}_+, 1 - E \in \mathcal{S}_+\} = \{E \in \mathcal{M}_d(\mathbb{C}) \mid 0 \leq_{\mathcal{S}_+} E \leq_{\mathcal{S}_+} 1\}, \quad (55)$$

describes a situation where all possible physical measurements are implementable, giving the strongest possible distinguishability (bias) between two quantum states,^{16,21} quantified by their trace distance. M_+ generates the cone $\mathcal{C}_{M_+} = \mathcal{S}_+ \subset \mathcal{V}^*$, so that the cone containing quantum states is $\mathcal{C} = \mathcal{S}_+ \subset \mathcal{V}$. Here, the base $\mathcal{B}_+ = \mathcal{S}_+ \cap H_1$ exactly equals the set of *all* physical quantum states, cf., Sec. II. The last expression in (55) shows $M_+ = \tilde{M}_+$ (cf., (50)), so that Theorem 14 gives the equality

$$\frac{1}{2} \|\rho_1 - \rho_2\|_1 = \inf \{\text{tr}[P_+] \mid \rho_1 - \rho_2 = P_+ - P_-, P_{\pm} \in \mathcal{S}_+\} = \sup_{0 \leq E \leq 1} \text{tr}[E(\rho_1 - \rho_2)]$$

of two well-known expressions for the trace distance between the quantum states ρ_1 and ρ_2 . Also, as $M \subseteq M_+$ for any other set M of physically implementable measurement operators,

$$\|\rho_1 - \rho_2\|_{(M)} \leq \|\rho_1 - \rho_2\|_1.$$

The capability of implementing all *separable measurements* on an n -partite quantum system corresponds to

$$M_{\text{SEP}} := \left\{ \sum_{k=1}^L E_k^{(1)} \otimes \cdots \otimes E_k^{(n)} \mid E_k^{(j)} \in \mathcal{S}_+, L \leq K, \sum_{k=1}^K E_k^{(1)} \otimes \cdots \otimes E_k^{(n)} = 1 \right\}, \quad (56)$$

all *PPT measurements* (see also Ref. 27) to

$$M_{\text{PPT}^+} := \left\{ E \in \mathcal{M}_{d_1 d_2 \dots d_n}(\mathbb{C}) \mid \forall I \subseteq \{1, \dots, n\} : \left(\bigotimes_{i \in I} T_i \otimes \bigotimes_{i \notin I} \text{id}_i \right) E \in M_+ \right\}, \quad (57)$$

where the last condition means PPT implementability with respect to any bipartition. It is easy to see that $M_{\text{SEP}} = \tilde{M}_{\text{SEP}}$ and $M_{\text{PPT}^+} = \tilde{M}_{\text{PPT}^+}$ (see Fig. 1), so that (54) holds with equality. The two classes (56) and (57) derive their importance from the fact that they are closer than M_+ to the set of 2-outcome measurements that can be implemented by local quantum operations and classical communication (LOCC-measurements, M_{LOCC}). This set is further diminished if communication

between the parties is not allowed for,

$$M_{\text{LO}} := \text{cl conv} \left\{ \sum_{(k_1, \dots, k_n) \in \mathcal{E}} E_{k_1}^{(1)} \otimes \dots \otimes E_{k_n}^{(n)} \mid \mathcal{E} \subseteq \{1, \dots, K\}^n, E_k^{(j)} \in \mathcal{S}_+, \sum_{k=1}^K E_k^{(j)} = 1 \forall j \right\}. \quad (58)$$

Therefore,

$$M_{\text{LO}} \subseteq M_{\text{LOCC}} \subseteq M_{\text{SEP}} \subseteq M_{\text{PPT}^+} \subseteq M_+, \quad (59)$$

and these inclusions lead to corresponding inequalities between the associated distinguishability norms. The cones generated by the first three sets are actually equal since for every $E \in M_{\text{SEP}}$ (56) one can easily find $E' \in M_{\text{LO}}$ (58) and $p > 0$ such that $E' = pE$, meaning that every separable measurement can be probabilistically implemented by local quantum operations. This gives

$$\mathcal{C}_{M_{\text{LO}}} = \mathcal{C}_{M_{\text{LOCC}}} = \mathcal{C}_{M_{\text{SEP}}} \subseteq \mathcal{C}_{M_{\text{PPT}^+}} \subseteq \mathcal{C}_{M_+} \quad \text{and} \quad \tilde{M}_{\text{LO}} = \tilde{M}_{\text{LOCC}} = \tilde{M}_{\text{SEP}} \subseteq \tilde{M}_{\text{PPT}^+} \subseteq \tilde{M}_+. \quad (60)$$

The upper two inclusions in each of the chains in (60) and (59) are known to be strict, at least in large enough dimensions.¹⁷ This is, however, not clear for the two lower inclusions in (59); it is known that, on a 3×3 -dimensional quantum system, separable measurements with 9 outcomes are strictly more powerful than LOCC-measurements,⁶ and although one may conjecture the same for the 2-outcome measurements in (59), to the best of our knowledge this has not been established. If, for example, the inclusion $M_{\text{LOCC}} \subseteq M_{\text{SEP}}$ were strict (see also Fig. 1), then we could find quantum states ρ_1, ρ_2 whose LOCC-distance is strictly smaller than their distance under the corresponding base norm, $\|\rho_1 - \rho_2\|_{(M_{\text{LOCC}})} < \|\rho_1 - \rho_2\|_{(\mathcal{C}_{M_{\text{LOCC}}})^*} = \|\rho_1 - \rho_2\|_{(M_{\text{SEP}})}$ (i.e., strict inequality in (54) and in Theorem 14).

Another set often used to approximate M_{LOCC} in a bipartite setting is $M_{\text{PPT}} := \{E | E^{T_1} \in M_+\} = (M_+)^{T_1}$ (see (20) for notation). This is neither a subset nor a superset of the physically implementable measurements M_+ , but rather a superset of M_{PPT^+} (cf., (59)). Nevertheless, it is often easier to handle in practice, and the theorems in this section apply to such “unphysical” sets of measurements as well. We will further discuss these approximations to M_{LOCC} and relations with Hilbert’s projective metric below Corollary 15 and in the contraction example below Lemma 17.

Note that for $M = M_{\text{SEP}}, M_{\text{LOCC}}, M_{\text{LO}}$ it is hard to express the corresponding cone $\mathcal{C} = (\mathcal{C}_{M_{\text{SEP}}})^*$ in an explicit form, as would be desirable in order to compute the corresponding base norm. But due to $\mathcal{C} \supseteq \mathcal{S}_+ \supset \mathcal{B}_+$ it is at least guaranteed that every physical state is an element of \mathcal{C} . For the other classes of measurements, however, the cones containing the states can be expressed explicitly: for M_{PPT} one has $\mathcal{C} = (\mathcal{S}_+)^{T_1}$, and for $M_{\text{PPT}^+} = \bigcap_{I \subseteq [n]} (M_+)^{T_I}$ (cf., (57)) it is $\mathcal{C} = \text{conv}(\bigcup_{I \subseteq [n]} (\mathcal{S}_+)^{T_I})$, which as the convex hull of convex sets is easily written down explicitly.

Using the duality between distinguishability and base norms from Theorem 14, Proposition 7 translates most directly to the present context of distinguishability measures and bounds them by Hilbert’s projective metric,

Corollary 15 (Distinguishability norm vs. Hilbert’s projective metric): For a finite-dimensional vector space \mathcal{V} and a distinguished element $e \in \mathcal{V}^*$, let $M \subset \mathcal{V}^*$ be a closed convex set with non-empty interior and satisfying (47); then M induces a distinguishability norm $\|\cdot\|_{(M)}$ on \mathcal{V} via (49), generates a proper cone $\mathcal{C}_M \subset \mathcal{V}^*$ and induces a proper cone $\mathcal{C} := (\mathcal{C}_M)^* \subset \mathcal{V}$ with base $\mathcal{B} := \mathcal{C} \cap H_e$. Let $b_1, b_2 \in \mathcal{B}$. Then,

$$\frac{1}{2} \|b_1 - b_2\|_{(M)} = \sup_{E \in M} \langle E, b_1 - b_2 \rangle \leq \frac{(\sup_{\mathcal{C}}(b_1/b_2) - 1)(1 - \inf_{\mathcal{C}}(b_1/b_2))}{\sup_{\mathcal{C}}(b_1/b_2) - \inf_{\mathcal{C}}(b_1/b_2)}, \quad (61)$$

$$\leq \frac{1}{1 + \inf_{\mathcal{C}}(b_1/b_2)} - \frac{1}{1 + \sup_{\mathcal{C}}(b_1/b_2)}, \quad (62)$$

$$\leq \tanh \frac{\mathfrak{h}_{\mathcal{C}}(b_1, b_2)}{4}. \quad (63)$$

Proof: This is an immediate consequence of Theorem 14 and the inequalities (28)–(30); cf., also Proposition 7. \square

Remark: As the lhs, the rhs in the chain of inequalities in Corollary 15 can likewise be written directly in terms of M ; with Eqs. (5) and (8):

$$\mathfrak{h}_C(b_1, b_2) = \mathfrak{h}_{(C_M)^*}(b_1, b_2) = \ln \sup_{E, F \in M} \frac{\langle E, b_1 \rangle \langle F, b_2 \rangle}{\langle E, b_2 \rangle \langle F, b_1 \rangle},$$

where in the context of quantum theory the last expression contains *ratios* of measurement probabilities.

Translating Corollary 15 into the quantum information context, Hilbert's projective metric yields a bound on the maximal bias in distinguishing two quantum states by a given set M of implementable measurements. If, for instance, all physical measurements are implementable ($M = M_+$, so $C = S_+$), one gets that the trace distance between two states $\rho_1, \rho_2 \in \mathcal{B}_+$ is upper bounded as follows:

$$\frac{1}{2} \|\rho_1 - \rho_2\|_1 \leq \tanh \frac{\mathfrak{h}_{S_+}(\rho_1, \rho_2)}{4}. \quad (64)$$

For M_+ and for other sets of measurements we will now examine such bounds in a concrete example.

Example (“data hiding”¹³). On a bipartite $d \times d$ -dimensional quantum system, consider the task of distinguishing the two Werner states $\rho_i = p_i \sigma_+ + (1 - p_i) \sigma_-$ (Ref. 38), $i = 1, 2$, where $\sigma_{\pm} = (1 \pm F)/d(d \pm 1)$ are the (anti)symmetric states, $F = \sum_{i,j} |ij\rangle\langle ji|$, and $0 \leq p_2 \leq p_1 \leq 1$. One can compute $\|\rho_1 - \rho_2\|_1 = 2(p_1 - p_2)$, $\|\rho_1 - \rho_2\|_{(M_{\text{PPT}})} = 4(p_1 - p_2)/d$ and $\|\rho_1 - \rho_2\|_{(M_{\text{PPT}^+})} = 4(p_1 - p_2)/(d + 1)$,²⁷ which are all upper bounds on $\|\rho_1 - \rho_2\|_{(M_{\text{LOCC}})}$ by (59). This enables “data hiding,”¹³ as the bias in distinguishing ρ_1 versus ρ_2 in a LOCC-measurement is smaller by a factor of order d than the best bias under all quantum measurements (in fact, $\|\rho_1 - \rho_2\|_{(M_{\text{LOCC}})} = 4(p_1 - p_2)/(d + 1)$).^{13,26}

Comparing these norms to the Hilbert metric bounds of Corollary 15, note first that $\mathfrak{h}_{S_{\text{PPT}}}(\rho_1, \rho_2)$ is not defined for $p_2 < 1/2$ since $\rho_2 \notin S_{\text{PPT}}$, whereas norm distances depend only on the difference $\rho_1 - \rho_2$. For the other cones,

$$\sup_{S_+}(\rho_1/\rho_2) = \sup_{(S_{\text{PPT}^+})^*}(\rho_1/\rho_2) = \frac{p_1}{p_2}, \quad \inf_{S_+}(\rho_1/\rho_2) = \frac{1 - p_1}{1 - p_2}, \quad \inf_{(S_{\text{PPT}^+})^*}(\rho_1/\rho_2) = \frac{d + 1 - 2p_1}{d + 1 - 2p_2}$$

(note $(S_+)^* = S_+$ and $(S_{\text{PPT}})^* = S_{\text{PPT}}$, whereas $(S_{\text{PPT}^+})^* = (S_+ \cap S_{\text{PPT}})^* = \text{conv}(S_+ \cup S_{\text{PPT}}) \supsetneq S_{\text{PPT}^+}$). So, (63) from Corollary 15 gives, for instance, for M_{PPT^+} , the upper bound (for large d)

$$\frac{1}{2} \|\rho_1 - \rho_2\|_{(M_{\text{PPT}^+})} \leq \tanh \frac{\mathfrak{h}_{(S_{\text{PPT}^+})^*}(\rho_1, \rho_2)}{4} = 1 - \frac{2}{1 + \sqrt{p_1/p_2}} + \mathcal{O}\left(\frac{1}{d}\right).$$

This bound by Hilbert's projective metric does not yield the $1/d$ behavior required for data hiding.

Significantly stronger bounds can be obtained for this example if one expresses them directly in terms of the sup and inf from above. For example, employing (61) gives upper bounds $\|\rho_1 - \rho_2\|_{(M_{\text{PPT}^+})} \leq 4(p_1 - p_2)/(d + 1)$ and $\|\rho_1 - \rho_2\|_1 \leq 2(p_1 - p_2)$, both of which coincide with the actual values and certify the possibility of data hiding. Tightness of the Hilbert metric bound (63) is lost in the arithmetic-geometric-mean inequality (26).

We will now translate Corollary 9 from the base norm language into a contractivity result for distinguishability norms. In general, by Theorem 14, a distinguishability norm is merely upper bounded by the corresponding base norm; but to obtain a consistent chain of inequalities, one needs equality in one place and this explains the condition $M = \tilde{M}$ in Proposition 16(b).

After formulating this contractivity result, we will state in Lemma 17 a few implications and equivalences regarding maps and their duals which allow for alternative formulations of the conditions in Proposition 16. Note that, in the quantum context, the process of measuring a quantum

system after the action of a quantum operation, expressed as $\langle E', T(\rho) \rangle$, can be described equivalently as evolution of the measurement operator under the dual map, since $\langle E', T(\rho) \rangle = \langle T^*(E'), \rho \rangle$ for all ρ and E' (“Heisenberg picture”). Hence the occurrence of T^* acting on measurement operators associated with the output space in the following.

Proposition 16 (Distinguishability norm contraction): For finite-dimensional vector spaces $\mathcal{V}, \mathcal{V}'$ and distinguished elements $e \in \mathcal{V}^*, e' \in \mathcal{V}'^*$ in their duals, let $M \subset \mathcal{V}^*$ and $M' \subset \mathcal{V}'^*$ be closed convex sets with non-empty interior and satisfying (47); they then generate proper cones $\mathcal{C}_M \subset \mathcal{V}^*, \mathcal{C}_{M'} \subset \mathcal{V}'^*$ and induce proper cones $\mathcal{C} := (\mathcal{C}_M)^* \subset \mathcal{V}, \mathcal{C}' := (\mathcal{C}_{M'})^* \subset \mathcal{V}'$ with bases $\mathcal{B} := \mathcal{C} \cap H_e, \mathcal{B}' := \mathcal{C}' \cap H_{e'}$. Let $T : \mathcal{V} \rightarrow \mathcal{V}'$ be a linear map. Then the following hold for all $v_1, v_2 \in \mathcal{V}$ with $\langle e, v_1 \rangle = \langle e, v_2 \rangle$:

- (a) If $T^*(M') \subseteq M$ and $T^*(e') = e$, then $\|T(v_1) - T(v_2)\|_{(M')} \leq \|v_1 - v_2\|_{(M)}$.
- (b) If T is base-preserving (i.e., $T(\mathcal{B}) \subseteq \mathcal{B}'$) and $M = \tilde{M}$, then

$$\|T(v_1) - T(v_2)\|_{(M')} \leq \|v_1 - v_2\|_{(M)} \tanh \frac{\Delta(T)}{4}. \quad (65)$$

Proof: For (a), note that

$$\begin{aligned} \|T(v_1 - v_2)\|_{(M')} &= \sup_{E' \in M'} \langle 2E' - e', T(v_1 - v_2) \rangle = \sup_{E' \in M'} \langle 2T^*(E') - T^*(e'), v_1 - v_2 \rangle \\ &= \sup_{E \in T^*(M')} \langle 2E - e, v_1 - v_2 \rangle \leq \sup_{E \in M} \langle 2E - e, v_1 - v_2 \rangle = \|v_1 - v_2\|_{(M)}. \end{aligned}$$

For (b), use $\|\cdot\|_{(M')} \leq \|\cdot\|_{\mathcal{B}'}$ and $\|\cdot\|_{(M)} = \|\cdot\|_{(\tilde{M})} = \|\cdot\|_{\mathcal{B}}$ from Theorem 14 and, as T is base-preserving, $\|T(v_1) - T(v_2)\|_{\mathcal{B}'} \leq \|v_1 - v_2\|_{\mathcal{B}} \tanh[\Delta(T)/4]$ from Corollary 9. \square

Remark: One might conjecture that (65) holds even under the (weaker) assumptions of Proposition 16(a); this, however, is not true in general (not even in the case $\mathcal{V} = \mathcal{V}', M = M'$), as one can find explicit examples where $\Delta(T) < \infty$ and nevertheless the best contraction coefficient in Proposition 16(a) is 1.

Lemma 17 (Maps and dual maps): Under the conditions of Proposition 16, the following hold:

- (a) T is cone-preserving (i.e., $T(\mathcal{C}) \subseteq \mathcal{C}'$) if and only if its dual $T^* : \mathcal{V}'^* \rightarrow \mathcal{V}^*$ is cone-preserving (i.e., $T^*(\mathcal{C}_{M'}) \subseteq \mathcal{C}_M$).
- (b) If T (or T^*) is cone-preserving, then T and T^* have equal projective diameter, i.e., $\Delta(T) = \Delta(T^*)$.
- (c) $T^*(e') = e \Leftrightarrow T(H_{e'}) \subseteq H_e \Leftrightarrow \forall v \in \mathcal{V} : \langle e, v \rangle = \langle e', T(v) \rangle$.
- (d) T is base-preserving (i.e., $T(\mathcal{B}) \subseteq \mathcal{B}'$) iff $T^*(\mathcal{C}_{M'}) \subseteq \mathcal{C}_M$ and $T^*(e') = e$.
- (e) $T^*(M') \subseteq M \Rightarrow T^*(\mathcal{C}_{M'}) \subseteq \mathcal{C}_M$ (i.e., T^* and T are cone-preserving).
- (f) T is base-preserving $\Rightarrow T^*(M') \subseteq \tilde{M}$, where \tilde{M} is defined in (50).

Proof: (a), (e), and (f) follow from the definitions. (c) and (d) hold since $e \in (\mathcal{C}_M)^\circ$ (see proof of Theorem 14), so that H_e and \mathcal{B} span all of \mathcal{V} . (b) follows easily by writing down the claim using the defining Eqs. (11), (8) and (5) and by noting that the suprema from (5) and (11) can be interchanged; proper care can also be taken of cases where denominators become 0. \square

In quantum information theory, when sets $M \subset \mathcal{V}^*$ and $M' \subset \mathcal{V}'^*$ corresponding to implementable 2-outcome measurements are fixed, a given general quantum channel T might not satisfy the conditions of Proposition 16(a) or (b). However, in many interesting situations it does, and we will now describe some of them, thereby providing a physical interpretation of Proposition 16 (see also previous examples in this section).

If M and M' correspond to the set of all physically possible measurements, i.e., $M, M' = M_+$, then $\mathcal{C}, \mathcal{C}' = \mathcal{S}_+$, so any physically implementable quantum channel T obeys the conditions of

Proposition 16(a) and (b). And when applied to quantum states $\rho_1, \rho_2 \in \mathcal{B}_+$, Proposition 16(a) just gives the well-known trace-norm contraction,³³ whereas (b) yields a possibly non-trivial contraction coefficient,

$$\|T(\rho_1) - T(\rho_2)\|_1 \leq \|\rho_1 - \rho_2\|_1 \tanh \frac{\Delta(T)}{4},$$

cf. also (38). This has the interpretation that the maximal bias in distinguishing ρ_1 and ρ_2 decreases by at least a factor of $\tanh[\Delta(T)/4]$ under the application of the quantum channel T .

The condition $T^*(M') \subseteq M$ also holds (i) for $M, M' = M_{\text{SEP}}$ sets of separable measurements (56) and separable superoperators T ,³¹ (ii) for sets of PPT measurements M_{PPT^+} (57) and positive PPT-preserving operations T (i.e., $T(\rho^{T_i})^{T_i} \in \mathcal{S}_+$ for any $\rho \in \mathcal{S}_+$ and for partial transposition T_i with respect to any bipartition $I \subseteq \{1, \dots, n\}$), and (iii) for the (unphysical) sets of M_{PPT} measurements and PPT operations (i.e., $T(\rho^{T_i})^{T_i} \in \mathcal{S}_+$ for any $\rho \in \mathcal{S}_+$). As $M = \tilde{M}$ in all three cases (see earlier in this section), if T is furthermore trace-preserving then Proposition 16(b) applies. For the frequently considered case of the PPT-distance, this reads

$$\begin{aligned} \|T(\rho_1) - T(\rho_2)\|_{(M_{\text{PPT}})} &= \|(T(\rho_1 - \rho_2))^{T_1}\|_1 \\ &\leq \|(\rho_1 - \rho_2)^{T_1}\|_1 \tanh[\Delta_{\text{SPT}}(T)/4] \\ &= \|\rho_1 - \rho_2\|_{(M_{\text{PPT}})} \tanh[\Delta_{\text{SPT}}(T)/4]. \end{aligned}$$

In Appendix B we compare $\Delta_{\mathcal{S}_+}(T)$ and $\Delta_{\text{SPT}}(T)$ for a depolarizing channel.

For M and M' corresponding to the set of LOCC measurements and for a quantum operation T implementable by LOCC, one has $T^*(M_{\text{LOCC}}) \subseteq M_{\text{LOCC}}$ from the remark on the Heisenberg picture preceding Proposition 16. Equation (65) is not guaranteed to hold for this case as possibly $M_{\text{LOCC}} \neq \tilde{M}_{\text{LOCC}}$. But Proposition 16(a) yields non-strict contraction for a trace-preserving LOCC-operation T ,

$$\|T(\rho_1) - T(\rho_2)\|_{(M_{\text{LOCC}})} \leq \|\rho_1 - \rho_2\|_{(M_{\text{LOCC}})},$$

meaning that the LOCC-distinguishability cannot increase under the application of a LOCC-channel.

VI. FIDELITY AND CHERNOFF BOUND INEQUALITIES

Another very popular distinguishability measure in quantum information theory is the so-called *fidelity*,^{15,28,35} which can be seen as a generalization of the overlap of pure quantum states to mixed states. For two density matrices ρ_1, ρ_2 , i.e., $\rho_1, \rho_2 \in \mathcal{M}_d(\mathbb{C})$ positive semidefinite with $\text{tr}[\rho_1] = \text{tr}[\rho_2] = 1$, the fidelity is defined as

$$F(\rho_1, \rho_2) := \text{tr} \left[\sqrt{\rho_1^{1/2} \rho_2 \rho_1^{1/2}} \right]. \quad (66)$$

It bounds the trace distance through the well-known inequality²⁸

$$1 - F(\rho_1, \rho_2) \leq \frac{1}{2} \|\rho_1 - \rho_2\|_1 \leq \sqrt{1 - F(\rho_1, \rho_2)^2}, \quad (67)$$

and we will in the following proposition relate the fidelity to Hilbert's projective metric on the cone \mathcal{S}_+ of positive semidefinite matrices. In fact, we will show that the upper bound in (67) fits in between both sides of the above established inequality (64),

Proposition 18 (Fidelity vs. Hilbert's projective metric): Let $\rho_1, \rho_2 \in \mathcal{M}_d(\mathbb{C})$ be two density matrices, and denote by \mathcal{S}_+ the cone of positive semidefinite matrices in $\mathcal{M}_d(\mathbb{C})$. Then,

$$\sqrt{1 - F(\rho_1, \rho_2)^2} \leq \tanh \frac{\mathfrak{h}_{\mathcal{S}_+}(\rho_1, \rho_2)}{4}. \quad (68)$$

Proof: Using $1 - \tanh^2 x = 1/\cosh^2 x$, the claim (68) is equivalent to

$$1 \leq \cosh[\mathfrak{h}_{\mathcal{S}_+}(\rho_1, \rho_2)/4] F(\rho_1, \rho_2). \quad (69)$$

Now, as is well-known,^{15,28} there exists a POVM $(E_i)_{i=1}^n$ (i.e., $E_i \in \mathcal{S}_+$, $\sum_{i=1}^n E_i = 1$) such that the numbers $p_i := \text{tr}[E_i \rho_1]$ and $q_i := \text{tr}[E_i \rho_2]$ satisfy

$$F(\rho_1, \rho_2) = \sum_{i=1}^n \sqrt{p_i q_i}. \quad (70)$$

The rhs is the so-called *classical fidelity* between the probability distributions induced by $(E_i)_{i=1}^n$ on ρ_1 and ρ_2 . With such POVM elements E_i , one has by definitions (8) and (5),

$$\mathfrak{h}_{\mathcal{S}_+}(\rho_1, \rho_2) = \ln \sup_{E, F \in \mathcal{S}_+} \frac{\text{tr}[E \rho_1] \text{tr}[F \rho_2]}{\text{tr}[E \rho_2] \text{tr}[F \rho_1]} \quad (71)$$

$$\begin{aligned} &\geq \ln \sup_{1 \leq i, j \leq n} \frac{\text{tr}[E_i \rho_1] \text{tr}[E_j \rho_2]}{\text{tr}[E_i \rho_2] \text{tr}[E_j \rho_1]} \\ &= \ln \left[\sup_i \frac{p_i}{q_i} \sup_j \frac{q_j}{p_j} \right] = \ln(M/m), \end{aligned} \quad (72)$$

where $M := \sup_i (p_i/q_i)$, $m := \inf_i (p_i/q_i)$ (defining $x/0 := \infty$ for $x > 0$, and omitting indices i with $p_i = q_i = 0$ in the \sup_i and \inf_i). Comparing this to (69) and using $\cosh x = (e^x + e^{-x})/2$ and (70), we are therefore done if we can show

$$1 \leq \cosh \left[\frac{1}{4} \ln \frac{M}{m} \right] F(\rho_1, \rho_2) = \frac{1}{2} \left[\left(\frac{M}{m} \right)^{1/4} + \left(\frac{m}{M} \right)^{1/4} \right] \sum_{i=1}^n \sqrt{p_i q_i}. \quad (73)$$

We begin by showing that, for each $i = 1, \dots, n$ separately,

$$\left[\left(\frac{M}{m} \right)^{1/4} + \left(\frac{m}{M} \right)^{1/4} \right] \sqrt{p_i q_i} \geq \left(\frac{1}{Mm} \right)^{1/4} p_i + (Mm)^{1/4} q_i. \quad (74)$$

For $p_i = q_i = 0$, this statement is trivial. If $p_i > 0 = q_i$ then $M = \infty$, so $\mathfrak{h}_{\mathcal{S}_+}(\rho_1, \rho_2) = \infty$ by (72) and (68) holds trivially; similarly for $q_i > 0 = p_i$. In all other cases, divide both sides by $\sqrt{p_i q_i}$ and set $x := \sqrt{p_i/q_i} \in [\sqrt{m}, \sqrt{M}]$. Then, (74) follows if

$$\left[\left(\frac{M}{m} \right)^{1/4} + \left(\frac{m}{M} \right)^{1/4} \right] \geq \left(\frac{1}{Mm} \right)^{1/4} x + (Mm)^{1/4} \frac{1}{x}$$

holds for all x with $\sqrt{m} \leq x \leq \sqrt{M}$. But this is clear since it holds with equality at the boundary points $x = \sqrt{m}, \sqrt{M}$, and the right-hand side is a convex function of x while the left-hand side is constant.

(73) follows now by summing (74) over $i = 1, \dots, n$:

$$\begin{aligned} \frac{1}{2} \left[\left(\frac{M}{m} \right)^{1/4} + \left(\frac{m}{M} \right)^{1/4} \right] \sum_{i=1}^n \sqrt{p_i q_i} &\geq \frac{1}{2} \sum_{i=1}^n \left(\frac{1}{Mm} \right)^{1/4} p_i + (Mm)^{1/4} q_i \\ &= \frac{1}{2} \left[\left(\frac{1}{Mm} \right)^{1/4} + (Mm)^{1/4} \right] \geq 1, \end{aligned}$$

where in the second step we used $\sum_i p_i = \text{tr}[\sum_i E_i \rho_1] = \text{tr}[1 \rho_1] = 1$ and similarly $\sum_i q_i = 1$, and the last step follows as the sum of a non-negative number and its inverse is lower bounded by 2. \square

The important fact about the fidelity (66) used in the proof is the existence of a POVM $(E_i)_i$ such that (70) holds. In fact, it is even true that^{15,28}

$$F(\rho_1, \rho_2) = \min_{(E_i)_i \text{ POVM}} \sum_i \sqrt{\text{tr}[E_i \rho_1] \text{tr}[E_i \rho_2]}, \quad (75)$$

where the optimization is over all physically implementable POVMs $(E_i)_{i=1}^n$ and the minimum is attained.

One can generalize Proposition 18 and inequality (67) to more general measurement settings (e.g., with locality restrictions as in Sec. V) if one defines a generalized fidelity for these situations suitably, which we will now do. Let \mathbb{M} denote the set of all measurements (some of them possibly having $n > 2$ outcomes) that are implementable in a given physical situation;²⁷ i.e., the elements of \mathbb{M} are collections $(E_i)_{i=1}^n$ of operators E_i with $E_i \in \mathcal{V}^*$, $\sum_i E_i = e$ and $n \geq 1$, where \mathcal{V}^* is the dual of a finite-dimensional vector space \mathcal{V} equipped with a distinguished element $e \in \mathcal{V}^* \setminus \{0\}$; cf., Sec. V for related notation. The POVM elements $E \in M$ that can occur in 2-outcome POVMs $(E, e - E)$ are then obtained by grouping together the outcomes of any other allowed POVM and by mixing them classically and taking limits,

$$M := \text{cl conv} \left\{ \sum_{i \in \mathcal{E}} E_i \mid (E_i)_{i=1}^n \in \mathbb{M}, \mathcal{E} \subseteq \{1, \dots, n\} \right\}. \quad (76)$$

We require that M have non-empty interior and that $M \cap (-M) = \{0\}$; then the other conditions on M around (47) will hold automatically so that the usual physically reasonable setup of Sec. V applies. In particular, the cone $\mathcal{C} := (C_M)^* \subset \mathcal{V}$ is proper and it is exactly the set of all elements $c \in \mathcal{V}$ such that $\langle E_i, c \rangle \geq 0$ for all E_i that occur as elements of a POVM $(E_i)_{i=1}^n \in \mathbb{M}$. Define then the *generalized fidelity* $F_{\mathbb{M}}$ of $b_1, b_2 \in \mathcal{B} := \mathcal{C} \cap H_e$ as

$$F_{\mathbb{M}}(b_1, b_2) := \inf_{(E_i)_{i=1}^n \in \mathbb{M}} \sum_{i=1}^n \sqrt{\langle E_i, b_1 \rangle \langle E_i, b_2 \rangle}. \quad (77)$$

Note also that, when the set M is induced as above by a set \mathbb{M} of general POVM measurements, the distinguishability norm $\|v\|_{(M)}$ (49) of $v \in \mathcal{V}$ can be written directly in terms of \mathbb{M} :²⁷

$$\|v\|_{(M)} = \sup_{(E_i)_{i=1}^n \in \mathbb{M}} \sum_{i=1}^n |\langle E_i, v \rangle|. \quad (78)$$

Then the following generalization of Proposition 18 and inequality (67) holds:

Proposition 19 (Generalized fidelity vs. Hilbert's projective metric and distinguishability norm): As in the previous paragraphs, let \mathbb{M} be such that M in (76) has non-empty interior and satisfies $M \cap (-M) = \{0\}$. Then the following expressions are well-defined, and for $b_1, b_2 \in \mathcal{B}$ it holds that

$$1 - F_{\mathbb{M}}(b_1, b_2) \leq \frac{1}{2} \|b_1 - b_2\|_{(M)} \leq \sqrt{1 - F_{\mathbb{M}}(b_1, b_2)^2} \leq \tanh \frac{\mathfrak{h}_{\mathcal{C}}(b_1, b_2)}{4}. \quad (79)$$

Proof: For the right inequality, everything goes through as in the proof of Proposition 18, except if the infimum in (77) is not attained; but in this case, a simple limit argument can replace the equality in (70). Note that the supremum used to define $\mathfrak{h}_{\mathcal{C}}(b_1, b_2)$ (the analogue of Eq. (71) above) now runs over $E, F \in \mathcal{C}^* = \mathcal{C}_M \supseteq M$, and that M (76) contains all POVM elements E_i that occur in any POVM $(E_i)_{i=1}^n \in \mathbb{M}$.

For the middle inequality, let $\|b_1 - b_2\|_{(M)} = \sum_{i=1}^n |\langle E_i, b_1 - b_2 \rangle|$ for some $(E_i)_{i=1}^n \in \mathbb{M}$, cf., (78); again, a simple limit argument can deal with the case when the supremum is not attained. Define $p_i := \langle E_i, b_1 \rangle$ and $q_i := \langle E_i, b_2 \rangle$, and without loss of generality the POVM elements E_i are ordered such that there exists $k \in \{1, \dots, n\}$ so that $p_i \geq q_i$ for $1 \leq i \leq k$, and $p_i \leq q_i$ for $k < i \leq n$. Define further $x := \sum_{i=1}^k p_i$ and $y := \sum_{i=1}^k q_i$. Thus $\|b_1 - b_2\|_{(M)} = (x - y) + ((1 - y) - (1 - x)) = 2(x - y)$, and so finally

$$\begin{aligned} \left(\frac{1}{2} \|b_1 - b_2\|_{(M)} \right)^2 + F_{\mathbb{M}}(b_1, b_2)^2 &\leq (x - y)^2 + \left(\sum_{i=1}^k \sqrt{p_i q_i} + \sum_{i=k+1}^n \sqrt{p_i q_i} \right)^2 \\ &\leq (x - y)^2 + \left(\left[\sum_{i=1}^k p_i \sum_{j=1}^k q_j \right]^{1/2} + \left[\sum_{i=k+1}^n p_i \sum_{j=k+1}^n q_j \right]^{1/2} \right)^2 \end{aligned}$$

$$\begin{aligned}
&= (x - y)^2 + \left(\sqrt{xy} + \sqrt{(1-x)(1-y)} \right)^2 \\
&= 1 - \left(\sqrt{x(1-x)} - \sqrt{y(1-y)} \right)^2 \leq 1,
\end{aligned}$$

where the second line uses the Cauchy-Schwarz inequality for each of the two sums.

To prove the leftmost inequality in (79), let $F_{\mathbb{M}}(b_1, b_2) = \sum_{i=1}^n \sqrt{p_i q_i}$ where p_i, q_i, k, x , and y are defined as above for an appropriate POVM $(E_i)_{i=1}^n \in \mathbb{M}$, again employing a limit argument if needed. Then,

$$\begin{aligned}
\frac{1}{2} \|b_1 - b_2\|_{(M)} + F_{\mathbb{M}}(b_1, b_2) &\geq (x - y) + \sum_{i=1}^k \sqrt{p_i q_i} + \sum_{i=k+1}^n \sqrt{p_i q_i} \\
&\geq \sum_{i=1}^k (p_i - q_i) + \sum_{i=1}^k \sqrt{q_i q_i} + \sum_{i=k+1}^n \sqrt{p_i p_i} \\
&= \sum_{i=1}^n p_i = 1. \quad \square
\end{aligned}$$

Hilbert's projective metric also gives an upper bound on the *Chernoff bound*, the asymptotic rate at which the error in symmetric quantum hypothesis testing vanishes.² Given either n copies of the quantum state ρ_1 or n copies of the state ρ_2 , with *a priori* probabilities π_1 and π_2 for either case, the minimal error in distinguishing the two situations is $P_{err}(n) = (1 - \|\pi_1 \rho_1^{\otimes n} - \pi_2 \rho_2^{\otimes n}\|_1)/2$ when allowed to perform any physically possible quantum measurement.^{16,21} If both π_1 and π_2 are non-zero, then $P_{err}(n)$ decays asymptotically as $P_{err}(n) \simeq e^{-\xi n}$ with the Chernoff rate $\xi = -\ln \min_{0 \leq s \leq 1} \text{tr}[\rho_1^s \rho_2^{1-s}]$ independent of π_1, π_2 .²

Proposition 20 (Chernoff bound vs. Hilbert distance): Let $\rho_1, \rho_2 \in \mathcal{M}_d(\mathbb{C})$ be two density matrices, and denote by S_+ the cone of positive semidefinite matrices in $\mathcal{M}_d(\mathbb{C})$. Then the Chernoff bound $\xi = -\ln \min_{0 \leq s \leq 1} \text{tr}[\rho_1^s \rho_2^{1-s}]$ is upper bounded via

$$\xi \leq \frac{\mathfrak{h}_{S_+}(\rho_1, \rho_2)}{2}. \quad (80)$$

Proof: In the limit of many copies, the exponential decay rate is independent of the (non-zero) prior probabilities;² therefore, set $\pi_1 = \pi_2 = 1/2$. Then, Corollary 15 in the form of inequality (64) and the additivity guaranteed by Corollary 2 give, for any $n \in \mathbb{N}$,

$$\begin{aligned}
P_{err}(n) &= \frac{1}{2} \left(1 - \frac{1}{2} \|\rho_1^{\otimes n} - \rho_2^{\otimes n}\|_1 \right) \geq \frac{1}{2} \left(1 - \tanh \frac{\mathfrak{h}_{S_+}(\rho_1^{\otimes n}, \rho_2^{\otimes n})}{4} \right) \\
&= \frac{1}{2} \left(1 - \tanh \frac{n \mathfrak{h}_{S_+}(\rho_1, \rho_2)}{4} \right) = \frac{e^{-n \mathfrak{h}/2}}{1 + e^{-n \mathfrak{h}/2}},
\end{aligned}$$

where we abbreviated $\mathfrak{h} := \mathfrak{h}_{S_+}(\rho_1, \rho_2)$. Finally,

$$\xi \equiv -\lim_{n \rightarrow \infty} \frac{1}{n} \ln P_{err}(n) \leq \lim_{n \rightarrow \infty} \left[-\frac{1}{n} \ln e^{-n \mathfrak{h}/2} + \frac{1}{n} \ln (1 + e^{-n \mathfrak{h}/2}) \right] = \frac{\mathfrak{h}}{2}. \quad \square$$

Remark: We conjecture even the following strengthening of Propositions 18 and 20:

$$\sqrt{1 - \left(\min_{0 \leq s \leq 1} \text{tr}[\rho_1^s \rho_2^{1-s}] \right)^2} \leq \tanh \frac{\mathfrak{h}_{S_+}(\rho_1, \rho_2)}{4}.$$

VII. OPERATIONAL INTERPRETATION

Birkhoff's theorem (Theorem 4) implies that the distance of two quantum states with respect to Hilbert's projective metric in the positive semidefinite cone \mathcal{S}_+ does not increase upon the application of a quantum channel. This property is shared by many distance measures, e.g., the ones based on the trace-norm, the relative entropy, the fidelity, and the χ^2 -divergence.³⁴ In the following we show that for Hilbert's metric, however, a converse of this theorem can be stated: in essence, contractivity with respect to Hilbert's projective metric decides whether or not there exists a *probabilistic* quantum operation that maps a given pair of input states to a given pair of (potential) output states. Note, that Hilbert's metric can here decide even about the existence of a *completely positive* map,²⁸ whereas most other results in the context of Hilbert's metric are oblivious to whether maps are completely positive or merely positive (i.e., cone-preserving). Conditions for the existence of completely positive maps in a different but related setting were considered previously in Refs. 4, 11, and 25.

Consider two pairs of density matrices $\rho_1, \rho_2 \in \mathcal{M}_d(\mathbb{C})$ and $\rho'_1, \rho'_2 \in \mathcal{M}_{d'}(\mathbb{C})$. Then the existence of a positive linear map T that acts as $T(\rho_i) = p_i \rho'_i$ for some $p_i > 0$ implies some simple compatibility relations for the corresponding supports: loosely speaking, whenever there is an inclusion of the input supports, then the same inclusion has to hold for the supports of the outputs. More specifically, if such T exists then the following implications hold:

$$\begin{aligned} \text{supp}[\rho_1] \subseteq \text{supp}[\rho_2] &\Rightarrow \text{supp}[\rho'_1] \subseteq \text{supp}[\rho'_2], \\ \text{and } \text{supp}[\rho_1] \supseteq \text{supp}[\rho_2] &\Rightarrow \text{supp}[\rho'_1] \supseteq \text{supp}[\rho'_2]. \end{aligned} \quad (81)$$

If the supports of both pairs are compatible in the above sense, we can formulate the following equivalence:

Theorem 21 (Converse of Birkhoff's theorem): *Let $\rho_1, \rho_2 \in \mathcal{M}_d(\mathbb{C})$ and $\rho'_1, \rho'_2 \in \mathcal{M}_{d'}(\mathbb{C})$ be two pairs of density matrices which satisfy the compatibility relations in Eq. (81). Then, there exists a completely positive linear map $T : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_{d'}(\mathbb{C})$ that acts as $T(\rho_i) = p_i \rho'_i$ for some $p_i > 0$, if and only if*

$$\mathfrak{h}_{\mathcal{S}_+}(\rho_1, \rho_2) \geq \mathfrak{h}_{\mathcal{S}_+}(\rho'_1, \rho'_2). \quad (82)$$

Proof: The “only if” part is a consequence of Birkhoff's theorem (Theorem 4), but follows also from more elementary arguments: as T is positive, expression (6) gives $\sup(\rho_1/\rho_2) \geq \sup(T(\rho_1)/T(\rho_2)) = (p_1/p_2) \sup(\rho'_1/\rho'_2)$ and similarly for the indices $1 \leftrightarrow 2$ interchanged, so that (82) follows. For the “if” part let us first consider the case where $\text{supp}[\rho_1] \subseteq \text{supp}[\rho_2]$. The subsequent constructive proof closely follows Ref. 23.

Let $M := \sup(\rho_1/\rho_2)$, $m := \inf(\rho_1/\rho_2)$, and M', m' be defined analogously for ρ'_1, ρ'_2 . We assume that ρ_1 and ρ_2 are linearly independent (i.e., $M > m$) since the statement becomes trivial otherwise. The inclusions of the supports imply that $M, M' < \infty$, and Eq. (82) can be written as $M/m \geq M'/m'$. Thus, due to the projective nature of \mathfrak{h} , we can rescale one of the outputs, say ρ'_1 , with a strictly positive factor such that

$$M' \leq M \quad \text{and} \quad m' \geq m.$$

ρ'_1 may now have trace different from 1, but normalization can be accounted for by adjusting p_1 at the end. Define $u := M\rho_2 - \rho_1$, $v := \rho_1 - m\rho_2$, and a linear map T' on the span of ρ_1 and ρ_2 by $T'(\rho_i) := \rho'_i$. Then $T'(u), T'(v), u$ and v are all positive semidefinite by construction. Moreover, u and v have non-trivial kernels that cannot be contained in the kernel of ρ_2 since otherwise M and m would not be extremal (i.e., would be in conflict with $M = \inf\{\lambda | \lambda \rho_2 \geq \rho_1\}$ or $m = \sup\{\lambda | \rho_1 \geq \lambda \rho_2\}$). In other words, there are vectors $\psi, \phi \in \mathbb{C}^d$ such that $v|\psi\rangle = u|\phi\rangle = 0$ but $v|\phi\rangle, u|\psi\rangle \neq 0$. Using those, we can define a linear map on $\mathcal{M}_d(\mathbb{C})$ as

$$T(\rho) := \frac{\langle \psi | \rho | \psi \rangle}{\langle \psi | u | \psi \rangle} T'(u) + \frac{\langle \phi | \rho | \phi \rangle}{\langle \phi | v | \phi \rangle} T'(v).$$

The properties mentioned above make this map well-defined and completely positive.²⁸ Moreover, T coincides with T' on u and v and by linearity therefore also on ρ_1 and ρ_2 .

Clearly, the same argument applies to the case $\text{supp}[\rho_2] \subseteq \text{supp}[\rho_1]$ by interchanging indices $1 \leftrightarrow 2$. What remains is thus the case in which there is no inclusion in either direction for the supports of the inputs, so that Eq. (82) reads $\infty \geq \mathfrak{h}_{S^+}(\rho'_1, \rho'_2)$, which is always true. And indeed, we can in this case always construct a map with the requested properties since there are vectors $\psi, \phi \in \mathbb{C}^d$ such that $\rho_1|\psi\rangle = \rho_2|\phi\rangle = 0$ but $\rho_1|\phi\rangle, \rho_2|\psi\rangle \neq 0$. This suggests

$$T(\rho) := \langle \phi | \rho | \phi \rangle \rho'_1 + \langle \psi | \rho | \psi \rangle \rho'_2.$$

□

To conclude this discussion, we give an operational interpretation of this result. As in the theorem above, assume that for a given finite set of pairs of density matrices $\{(\rho_i, \rho'_i)\}$ there exists a completely positive linear map $T : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_{d'}(\mathbb{C})$ such that $T(\rho_i) = p_i \rho'_i$ for some $p_i > 0$. Then we can construct a new linear map $\tilde{T} : \mathcal{M}_d(\mathbb{C}) \rightarrow \mathcal{M}_{d'}(\mathbb{C}) \otimes \mathcal{M}_2(\mathbb{C})$ which is completely positive and trace-preserving and such that (i) it maps $\rho_i \mapsto \rho'_i$ conditioned on outcome “1” on the ancillary two-level system, and (ii) for any of the inputs ρ_i the outcome “1” is obtained with non-zero probability. More explicitly, this is obtained by

$$\tilde{T}(\rho) := cT(\rho) \otimes |1\rangle\langle 1| + B\rho B^\dagger \otimes |0\rangle\langle 0|,$$

where $c := \|T^*(1)\|_\infty^{-1}$ and $B := \sqrt{1 - cT^*(1)}$. Conversely, if a completely positive linear map \tilde{T} satisfying (i) and (ii) exists for a given set $\{(\rho_i, \rho'_i)\}$, then one can get a suitable map T by $T(\rho) := \langle 1 | \tilde{T}(\rho) | 1 \rangle$.

In other words, Theorem 21 shows that Hilbert’s projective metric provides a necessary and sufficient condition for the existence of a probabilistic quantum operation that maps $\rho_i \mapsto \rho'_i$ upon success. Note that the criterion (82) can be decided efficiently, for instance, by Proposition 1, as can the necessary condition (81).

VIII. CONCLUSION

We have introduced Hilbert’s projective metric into quantum information theory, where different convex sets and cones appear (such as the cones of positive semidefinite or of separable matrices), and where corresponding cone-preserving maps are ubiquitous (e.g., completely positive maps or LOCC operations). Hilbert’s projective metric, which is defined on any convex cone, is thus a natural tool to use in this context. We have found connections and applications to entanglement measures, via base norms and negativities, and to measures for statistical distinguishability of quantum states.

In particular, the projective diameter of a quantum channel yields contraction bounds for distinguishability measures and for entanglement measures under application of the channel. Such non-trivial contraction coefficients are hard to obtain by other means. For instance, whereas the second-largest eigenvalue of a channel determines its asymptotic contraction rates, the same is not true for its finite-time contraction behavior (albeit frequently assumed so). The projective diameter, however, yields valid contraction ratios even for the initial time.

These contraction results may sometimes be tools of more theoretical than practical interest, e.g., by being a guarantee for strict exponential contractivity. This is because, on the one hand, Hilbert’s projective metric $\mathfrak{h}_{\mathcal{C}}(a, b)$ is efficiently computable given an efficient description of \mathcal{C} by using Eq. (6). On the other hand, however, the definition of the projective diameter $\Delta(T)$ does not directly entail convex optimization: even though the maximization in Eq. (11) can be taken over the compact convex set $\mathcal{B} \times \mathcal{B}$ (with any base \mathcal{B} of \mathcal{C}), the function $\mathfrak{h}_{\mathcal{C}}$ is not jointly concave, as is intuitively apparent since $\mathfrak{h}_{\mathcal{C}}(a, b)$ grows when a, b approach the boundary of \mathcal{C} (see also Fig. 2(a)). In Appendices A and B we have seen examples where $\Delta(T)$ was exactly computable and other examples where this seemed not easy. Nevertheless, even non-trivial upper bounds on $\Delta(T)$ yield non-trivial contraction ratios and ensure immediate exponential convergence.

Besides these contractivity results, Hilbert's projective metric with respect to the positive semidefinite cone decides the possibility of extending a completely positive map, thereby yielding an operational interpretation.

ACKNOWLEDGMENTS

The authors thank M. A. Jivulescu and T. Heinosaari for valuable discussions. D.R. was supported by the European projects QUEVADIS and COQUIT. M.J.K. acknowledges financial support by the Niels Bohr International Academy. M.M.W. was supported by the Danish Research Council, F.N.U. and the Alfried Krupp von Bohlen und Halbach-Stiftung. M.M.W. is grateful to the Mittag-Leffler program, where part of the work has been carried out during fall 2010.

APPENDIX A: HILBERT'S PROJECTIVE METRIC FOR QUBITS

In this appendix we will, as an example, look at Hilbert's projective metric on the space associated with a two-level quantum system (qubit) and analyze how the projective diameter of qubit channels changes when choosing different cones (cf., discussion below Proposition 12). But before considering more general cones in the space \mathcal{V} of Hermitian 2×2 -matrices, we will specially examine Hilbert's metric associated with the positive semidefinite cone $\mathcal{S}_+ \subset \mathcal{V}$. The partial order induced by \mathcal{S}_+ is exactly the partial time-ordering of events $x = (x^0, x^1, x^2, x^3)$ in 4-dimensional Minkowski spacetime, which can be identified with \mathcal{V} via $x \mapsto \sum_{\mu} x^{\mu} \sigma_{\mu}$ where σ_0 and σ_i are the identity and Pauli matrices; Hilbert's projective metric has been considered in this situation before.²⁴ Furthermore, equipping a base of \mathcal{S}_+ (such as the set \mathcal{B}_+ of density matrices on a qubit) with Hilbert's projective metric gives the Beltrami-Klein model of projective geometry, in which the metric is usually written in terms of a *cross-ratio* of points, see Fig. 2(a).

Recall that in the Bloch sphere picture²⁸ each qubit state $\rho \in \mathcal{B}_+$ corresponds via $\rho = (1 + \vec{r} \cdot \vec{\sigma})/2$ to a point $\vec{r} \in \mathbb{R}^3$ in the unit sphere, $|\vec{r}| \leq 1$; we will freely identify ρ with \vec{r} and $\tau = (1 + \vec{t} \cdot \vec{\sigma})/2$ with \vec{t} , etc. Using expressions (6) and (7) and the fact that $\rho \leq_{\mathcal{S}_+} M\tau$ iff $M\tau - \rho$ has non-negative determinant and trace, one obtains explicitly (cf., also Ref. 24):

$$\mathfrak{h}_{\mathcal{S}_+}(\rho, \tau) = \ln \frac{1 - \vec{r} \cdot \vec{t} + \sqrt{(1 - \vec{r} \cdot \vec{t})^2 - (1 - \vec{r}^2)(1 - \vec{t}^2)}}{1 - \vec{r} \cdot \vec{t} - \sqrt{(1 - \vec{r} \cdot \vec{t})^2 - (1 - \vec{r}^2)(1 - \vec{t}^2)}}. \quad (\text{A1})$$

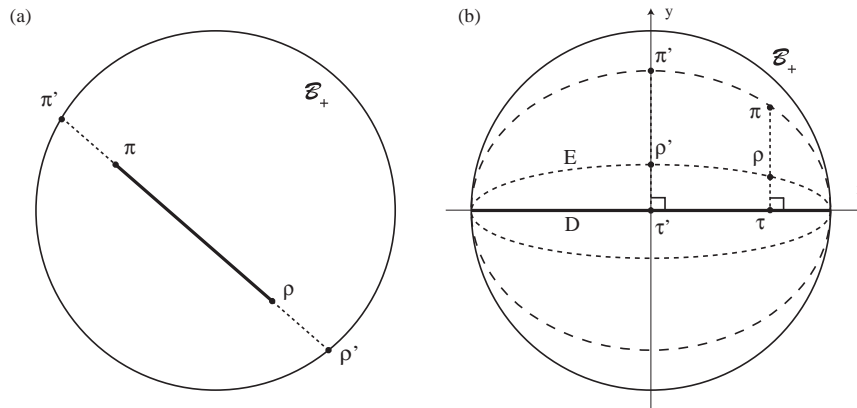


FIG. 2. (a) Hilbert's projective metric between two points $\pi, \rho \in \mathcal{B}_+$ of a base of \mathcal{S}_+ may be expressed as a logarithmic cross-ratio of Euclidean distances: $\mathfrak{h}_{\mathcal{S}_+}(\pi, \rho) = \ln(|\pi' - \rho|/|\rho' - \pi|/|\pi' - \pi|/|\rho' - \rho|)$.²³ (b) For $\rho = (x, y, 0)$, $\rho' = (0, y/\sqrt{1-x^2}, 0)$ and their projections $\tau = (x, 0, 0)$, $\tau' = (0, 0, 0)$ onto a diameter D of the Bloch sphere (here the x axis), one has $\mathfrak{h}_{\mathcal{S}_+}(\rho, \tau) = \mathfrak{h}_{\mathcal{S}_+}(\rho', \tau')$. Similarly, $\mathfrak{h}_{\mathcal{S}_+}(\pi, \tau) = \mathfrak{h}_{\mathcal{S}_+}(\pi', \tau')$, and the additivity of Hilbert's projective metric on lines yields $\mathfrak{h}_{\mathcal{S}_+}(\rho, \pi) = \mathfrak{h}_{\mathcal{S}_+}(\rho', \pi')$ in the geometric situation here. Note that the Euclidean distance, i.e., the trace distance,²⁸ is in general not preserved: $||\rho' - \pi'||_1 > ||\rho - \pi||_1$ if $\rho \neq \pi$ and $x \neq 0$.

Figure 2(b) illustrates that Hilbert's distance between any point ρ and its (Euclidean orthogonal) projection τ onto any diameter D of the Bloch sphere equals the distance between any other point ρ' on the ellipse E through ρ with major axis D and its projection τ' onto D ; this follows directly from (A1). In particular, for $\tau' = 1/2$ one has $\vec{t}' = 0$ and $\mathfrak{h}_{S_+}(\rho', 1/2) = \ln(1 + |\vec{r}'|)/(1 - |\vec{r}'|)$. This ellipse construction will be used below, as will the fact that Hilbert's projective metric is additive on lines, i.e., $\mathfrak{h}(\pi, \tau) = \mathfrak{h}(\pi, p\pi + q\tau) + \mathfrak{h}(p\pi + q\tau, \tau)$ for $p, q \geq 0$.²³ Note that all figures here show a 2-dimensional cross section through the Bloch sphere.

We will now consider positive linear and trace-preserving maps on qubits, using this geometric picture. Such a map T acts on the Bloch sphere representation of ρ as $T(\vec{r}) = \Lambda\vec{r} + \vec{v}$ with a matrix $\Lambda \in \mathbb{R}^{3 \times 3}$ and $\vec{v} \in \mathbb{R}^3$. Since unitary transformations, corresponding to $SO(3)$ rotations of the Bloch sphere, leave the qubit state space \mathcal{B}_+ invariant, the image $T(\mathcal{B}_+)$ of the Bloch sphere is an ellipsoid with semi-principal axes given by the singular values of Λ , shifted away from the origin by \vec{v} . Unital maps are exactly the ones with $\vec{v} = 0$.

As the trace distance between qubit states coincides with their Euclidean distance in the Bloch sphere picture,²⁸ Proposition 12 immediately gives the *trace-norm contraction coefficient* $\eta_1(T) := \eta_{\mathcal{B}_+}^b(T) = \|\Lambda\|_\infty$ (largest singular value of Λ). Recall from (11) that, similarly, the projective diameter $\Delta(T)$ is defined as the largest diameter of the image $T(\mathcal{B}_+)$, measured via Hilbert's projective metric \mathfrak{h}_{S_+} . $\Delta(T)$ is hard to express in terms of Λ and \vec{v} , but Corollary 9 proves $\tanh[\Delta(T)/4]$ to be an upper bound on the trace-norm contraction coefficient $\eta_1(T) = \|\Lambda\|_\infty$, and for maps on qubits we can actually characterize the cases of equality:

Proposition 22 (Trace-norm contraction vs. projective diameter for qubits): For a linear map $T : \mathcal{B}_+ \rightarrow \mathcal{B}_+$ on qubits, the inequality $\eta_1(T) \leq \tanh[\Delta(T)/4]$ holds with equality if and only if T is unital or constant (i.e., mapping \mathcal{B}_+ onto one point).

Proof: If T is unital, the image $T(\mathcal{B}_+)$ is an ellipsoid centered about the origin. In this symmetric situation, the largest Hilbert distance between any two points of this ellipsoid is the distance $\mathfrak{h}_{S_+}(\rho, \pi)$ between the two extremal points ρ and π of its major axis; this follows easily from the cross-ratio definition of Hilbert's projective metric (Fig. 2(a)), as this pair of points maximizes their Euclidean distance $\|\rho - \pi\|_1$ while at the same time minimizing the Euclidean distances $\|\rho - \rho'\|_1$ and $\|\pi - \pi'\|_1$ to the boundary. Thus,

$$\Delta(T) = \mathfrak{h}_{S_+}(\rho, \pi) = \mathfrak{h}_{S_+}(\rho, 1/2) + \mathfrak{h}_{S_+}(1/2, \pi) = 2 \ln \frac{1 + \|\Lambda\|_\infty}{1 - \|\Lambda\|_\infty},$$

and a little algebra yields $\tanh[\Delta(T)/4] = \|\Lambda\|_\infty = \eta_1(T)$. If T is constant, then $\eta_1(T) = \Delta(T) = 0$, so equality holds as well.

Conversely, if T is neither unital nor constant, denote by π and ρ the extremal points of the major axis of $T(\mathcal{B}_+)$. Then find a diameter D of the Bloch sphere that yields the construction from Fig. 2(b), i.e., choose D such that π and ρ have the same Euclidean orthogonal projection onto D . It is easy to see (e.g., by the cross-ratio) that centering π' and ρ' along their connecting line about the origin does not increase their Hilbert distance; i.e., denoting $\pi = (x, y, 0)$ in addition to the caption of Fig. 2(b) and defining $\pi'', \rho'' := (0, \pm(y' - y)/2\sqrt{1 - x^2}, 0)$, one has $\mathfrak{h}_{S_+}(\rho'', \pi'') \leq \mathfrak{h}_{S_+}(\rho', \pi') = \mathfrak{h}_{S_+}(\rho, \pi)$. Thus,

$$\tanh \frac{\mathfrak{h}_{S_+}(\pi, \rho)}{4} \geq \tanh \frac{\mathfrak{h}_{S_+}(\pi'', \rho'')}{4} = \frac{|y' - y|}{2\sqrt{1 - x^2}} = \frac{\|\pi - \rho\|_1}{2\sqrt{1 - x^2}} \geq \frac{\|\pi - \rho\|_1}{2}.$$

As T is not unital, at least one of the two transformations $(\pi, \rho) \rightarrow (\pi', \rho') \rightarrow (\pi'', \rho'')$ was not the identity, such that at least one of the two inequalities in the above chain is strict. This, together

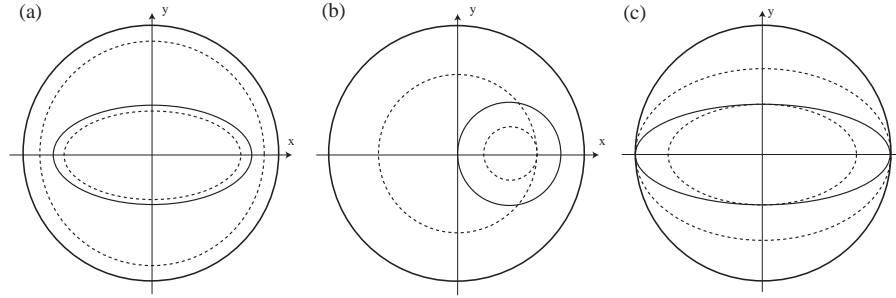


FIG. 3. The solid lines indicate the Bloch sphere \mathcal{B}_+ and its images $T(\mathcal{B}_+)$, whereas the dashed lines show the restricted cones and their images under T . (a) and (b) Spherically symmetric restriction of the Bloch sphere, (a) with a unital map, and (b) with a non-unital map. (c) Restriction of the Bloch sphere to an ellipsoid, with a unital channel that rotates and deforms the Bloch sphere.

with $\Delta(T) \geq \mathfrak{h}_{\mathcal{S}_+}(\pi, \rho)$ (by Definition 3) and $\eta_1(T) = \|\pi - \rho\|_1/2$ (by Proposition 12), yields $\tanh[\Delta(T)/4] > \eta_1(T)$. \square

Some more general cones can be conveniently parametrized in the Bloch representation: For a non-negative function $f(\hat{r})$ on unit vectors $|\hat{r}| = 1$ in \mathbb{R}^3 , the set

$$\mathcal{B}_f := \{\rho = (1 + \vec{r} \cdot \vec{\sigma})/2 \mid |\vec{r}| \leq f(\hat{r})\} \quad (\text{A2})$$

of normalized Hermitian matrices forms the base of a convex cone \mathcal{C}_f if \mathcal{B}_f is itself convex. $f \equiv 1$ gives the set of density matrices \mathcal{B}_+ and the positive semidefinite cone \mathcal{S}_+ , whereas $f \equiv c \in (0, \infty)$ yields the cone $\mathcal{C}_{f \equiv c}$ of all Hermitian 2×2 -matrices whose ratio of eigenvalues lies in a certain range. The defining equation (8) or, equivalently, the cross-ratio (Fig. 2(a)) allow for explicit computation of the Hilbert distance from the origin,

$$\mathfrak{h}_{\mathcal{C}_f}(\rho, 1/2) = \ln \frac{1 + |\vec{r}|/f(-\hat{r})}{1 - |\vec{r}|/f(\hat{r})}. \quad (\text{A3})$$

We can now analyze how the projective diameter of a map T changes when changing the cone (cf., discussion below Proposition 12). Of course, in order for the projective diameter to be well-defined, T has to preserve the cone in question. By looking at examples in which the cone \mathcal{S}_+ is being restricted to subcones, we find cases (a) where the diameter stays the same, (b) where it increases, and (c) where it decreases; see Fig. 3.

- (a) For any unital channel T , the projective diameter does not change when restricting \mathcal{S}_+ to a subcone \mathcal{C}_f with $f \equiv c \in (0, 1)$, i.e., when shrinking the cone spherically symmetrically. The ellipsoid $T(\mathcal{B}_f)$ is scaled down by a factor c compared to $T(\mathcal{B}_+)$, but, as (A3) already indicates, Hilbert distances depend only on *ratios* of Euclidean distances, so that $\Delta_{\mathcal{S}_+}(T) = \Delta_{\mathcal{C}_{f \equiv c}}(T)$.
- (b) Consider the channel T with $\Lambda = 1_3/3$ and $\vec{v} = (1/3, 0, 0)$, see Fig. 3(b). Restricting to the same subcone $\mathcal{C}_{f \equiv c}$ as in (a), T is cone-preserving if and only if $c \geq 1/2$. Clearly, the projective diameter $\Delta_{\mathcal{S}_+}(T)$ with respect to the cone \mathcal{S}_+ is finite as $T(\mathcal{B}_+)$ stays away from the boundary of \mathcal{B}_+ , whereas $\Delta_{\mathcal{C}_{f \equiv 1/2}}(T) = \infty$ as $T(\mathcal{B}_{f \equiv 1/2})$ touches the boundary of $\mathcal{B}_{f \equiv 1/2}$.
- (c) The unital channel T here rotates the Bloch sphere and shrinks it anisotropically: $\Lambda_{1,2} = 1$, $\Lambda_{2,1} = \Lambda_{3,3} = 1/2$, and $\Lambda_{i,j} = 0$ otherwise. Clearly, $\Delta_{\mathcal{S}_+}(T) = \infty$ as $T(\mathcal{B}_+)$ touches the boundary of \mathcal{B}_+ . But if one takes the restricted cone \mathcal{C} to be generated by an ellipsoidal base $\mathcal{B} \subset \mathcal{B}_+$ with major axis identical to the major axis of $T(\mathcal{B}_+)$ and with the other two principal axes slightly shortened, then $T(\mathcal{B})$ stays away from the boundary of \mathcal{B} , so that $\Delta_{\mathcal{C}}(T) < \infty$.

These examples show that the projective diameter is not monotonic under the restriction to subcones. Of course, more generally, the cones \mathcal{C} in the domain and \mathcal{C}' in the codomain do not have to coincide and can be varied independently. Then, monotonicity under the restriction of either \mathcal{C} or \mathcal{C}' holds as noted below Proposition 12.

APPENDIX B: PROJECTIVE DIAMETER OF DEPOLARIZING CHANNELS

Here, we compute the projective diameter for a well-known family of quantum channels, thereby also illustrating the contraction bounds from Sec. IV. We will first concentrate on quantities associated with the positive semidefinite cone \mathcal{S}_+ , and later comment on a bipartite scenario and the cone \mathcal{S}_{PPT} of PPT matrices.

A general depolarizing quantum channel on a d -dimensional system can be written as

$$T(\rho) = p\rho + (1-p)\text{tr}[\rho]\sigma, \quad (\text{B1})$$

with a density matrix σ (“fixed point”) and a probability parameter $p \in [0, 1]$. The trace-norm contraction coefficient of T , or any other norm contraction coefficient obtained by using the same norm in both the domain and codomain of T , is given by $\eta^b(T) = p$, as $\|T(\rho_1) - T(\rho_2)\| = p\|\rho_1 - \rho_2\|$ for all ρ_1, ρ_2 . Note that this contraction coefficient is independent of the fixed point σ . However, as we will see now, σ does influence the projective diameter $\Delta(T)$, from which upper bounds on the trace-norm contraction coefficient can be obtained.

To compute the projective diameter $\Delta(T)$ of T with respect to the positive semidefinite cone \mathcal{S}_+ , denote the eigenvalues of σ by $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_d$ with corresponding eigenvectors ψ_1, \dots, ψ_d (henceforth, assume $d \geq 2$). One can see that

$$M_{ij} := \sup (T(\psi_i)/T(\psi_j)) = 1 + \frac{p}{(1-p)\lambda_i} \quad \text{for } i \neq j,$$

as M_{ij} is the smallest number such that $M_{ij}T(\psi_j) - T(\psi_i) = (M_{ij} - 1)(1-p)\sigma + M_{ij}p\psi_j - p\psi_i$ is positive semidefinite, see Eq. (6). Maximizing only over the eigenstates of σ , one thus obtains the lower bound,

$$\Delta(T) \geq \mathfrak{h}_{\mathcal{S}_+}(T(\psi_1), T(\psi_2)) = \ln \left[\left(1 + \frac{p}{(1-p)\lambda_1} \right) \left(1 + \frac{p}{(1-p)\lambda_2} \right) \right]. \quad (\text{B2})$$

On the other hand $\sup(T(\rho_1)/T(\rho_2)) \leq M_{12}$ for any density matrices ρ_1, ρ_2 , since

$$M_{12}T(\rho_2) - T(\rho_1) = p(\sigma/\lambda_1 - \rho_1) + pM_{12}\rho_2 \geq p(1 - \rho_1) \geq 0,$$

so that, from the defining Eqs. (8) and (11),

$$\Delta(T) \leq \ln M_{12}^2 = \ln \left(1 + \frac{p}{(1-p)\lambda_1} \right)^2. \quad (\text{B3})$$

From these expressions it is clear that the projective diameter $\Delta(T)$ depends not solely on the depolarizing parameter p , but also on the spectrum of the fixed point σ . The lower and upper bounds (B2) and (B3) coincide if the lowest eigenvalue of σ is degenerate, for instance, in the case of depolarization towards the completely mixed state $\sigma = 1/d$. In any case, the upper bound on the trace-norm contraction coefficient $\eta_1(T) := \eta^b(T)$ obtained from Corollary 9 and (B3) is

$$\eta_1(T) \leq \tanh \frac{\Delta(T)}{4} \leq \frac{1}{1 + 2\lambda_1(1-p)/p}.$$

This is stronger than the trivial upper bound $\eta_1(T) \leq 1$, but weaker than the true value $\eta_1(T) = p$.

If the state space is bipartite, one can consider depolarization towards a separable quantum state σ (or towards any PPT state σ). This depolarizing map preserves then also the cone \mathcal{S}_{PPT} of PPT matrices. Since the positive semidefinite cone is related via partial transposition to $\mathcal{S}_{\text{PPT}} = (\mathcal{S}_+)^{T_1}$, it follows easily from the definition of the projective diameter, that the diameter with respect to \mathcal{S}_{PPT} of the depolarizing channel $T_{p,\sigma}$ from Eq. (B1) is equal to the diameter with respect to \mathcal{S}_+ of the channel $T_{p,\sigma^{T_1}}$ that effects depolarization towards the partially transposed state σ^{T_1} :

$$\Delta_{\mathcal{S}_{\text{PPT}}}(T_{p,\sigma}) = \Delta_{\mathcal{S}_+}(T_{p,\sigma^{T_1}}). \quad (\text{B4})$$

As an example, the Werner state $\sigma_q := q\sigma_+ + (1-q)\sigma_-$ on a $d \times d$ -dimensional system (for notation, see, the example, Corollary 15 below) is separable (and PPT) iff $1/2 \leq q \leq 1$, and its

partial transpose is

$$\sigma_q^{T_1} = 1 \left(\frac{q}{d(d+1)} + \frac{1-q}{d(d-1)} \right) + \Omega \left(\frac{q}{d+1} - \frac{1-q}{d-1} \right)$$

with the maximally entangled state $\Omega := \sum_{i,j} |ii\rangle\langle jj|/d = F^{T_1}/d$. Assume $d \geq 3$ such that the lowest eigenvalue $\lambda_1 = \min\{2q/d(d+1), 2(1-q)/d(d-1)\}$ of σ_q is always degenerate and the diameter $\Delta_{S_+}(T_{p,\sigma_q})$ is given by the rhs of (B3). Now, for $q \geq (d+1)/2d$ the lowest eigenvalue λ'_1 of $\sigma_q^{T_1}$ is degenerate as well and given by the first parentheses in the previous equation; thus, $\Delta_{S_{\text{PPT}}}(T_{p,\sigma_q})$ can be computed via (B4) and (B3), and one finds for $q > (d+1)/2d$ that, because of $\lambda_1 < \lambda'_1$, the diameter of T_{p,σ_q} is larger with respect to the cone S_+ than with respect to the cone S_{PPT} . For $q \in [1/2, (d+1)/2d)$ the lowest two eigenvalues λ'_1, λ'_2 of $\sigma_q^{T_1}$ are not degenerate; but the explicit lower bound (B2) on $\Delta_{S_+}(T_{p,\sigma_q^{T_1}}) = \Delta_{S_{\text{PPT}}}(T_{p,\sigma_q})$ is already sufficient to show that the ordering of both diameters is reversed for this range of q .

In conclusion, $\Delta_{S_+}(T_{p,\sigma_q}) < \Delta_{S_{\text{PPT}}}(T_{p,\sigma_q})$ for $q \in [1/2, (d+1)/2d)$, and $\Delta_{S_+}(T_{p,\sigma_q}) > \Delta_{S_{\text{PPT}}}(T_{p,\sigma_q})$ for $q \in ((d+1)/2d, 1]$, and equality holds for $q = (d+1)/2d$, i.e., when T is unital ($\sigma_q = \sigma_q^{T_1} = 1/d^2$).

APPENDIX C: OPTIMALITY OF BOUNDS AND CONTRACTION COEFFICIENTS

Here we show that the upper bounds given in Propositions 8 and 10 and in Corollary 9 are best possible in a specific sense. This also explains the appearance of the hyperbolic tangent in these statements when they are to be tight. As a consequence, Propositions 13 and 16 are optimal in the same sense. And a similar argument holds for the upper bounds in Proposition 7 and Corollary 15 (but cf., the remark, respectively, the example below each of the latter two statements).

First note that the Birkhoff-Hopf theorem (Theorem 4) guarantees that for *any* positive linear map T the contraction ratio $\tanh[\Delta(T)/4]$ is optimal when measuring distances by either Hilbert's projective metric or by the oscillation. As the qubit example in Appendix A (Proposition 22) already shows, this optimality for *any* map T does not hold for the negativity nor for the base norm contraction of Propositions 8 and 10. We can, however, demonstrate something weaker, namely, that for given proper cones $\mathcal{C}, \mathcal{C}'$ with bases $\mathcal{B}, \mathcal{B}'$ and for given diameter $\Delta \in (0, \infty)$ one can always find a base-preserving linear map $T : \mathcal{C} \rightarrow \mathcal{C}'$ with $\Delta(T) = \Delta$ and an element $v \in \mathcal{V}$ such that the contraction bounds in Propositions 8 and 10 are non-trivial and tightest possible, provided that the contraction factors are to depend on $\Delta(T)$ solely.

Before constructing such a map, we point out that in the proofs of both Propositions 8 and 10 the subtraction F is taken to be a linear combination of $T(b_1)$ and $T(b_2)$, while enforcing both terms in the representation $T(v) = (\lambda_1 T(b_1) - F) - (\lambda_2 T(b_2) - F)$ to be elements of the cone \mathcal{C}' . In the notation of the proofs, this allows an optimal F_{opt} which satisfies, as one can calculate,

$$\langle e', F \rangle \leq \langle e', F_{\text{opt}} \rangle = \frac{Mm\lambda_1 + \lambda_2 - m\lambda_1 - m\lambda_2}{M - m}. \quad (\text{C1})$$

Further maximization over an allowed range for m and M motivates their choice in the following construction:

To construct the desired map T , choose elements $b_1, b_2 \in \mathcal{B}$, $b'_1, b'_2 \in \mathcal{B}'$ of the bases with $\|b_1 - b_2\|_{\mathcal{B}} = \|b'_1 - b'_2\|_{\mathcal{B}'} = 2$ (see, e.g., beginning of the proof of Proposition 12), and for $0 \leq \mu_1 \leq \mu_2 \leq 1$ define $c'_i := (1 - \mu_i)b'_1 + \mu_i b'_2 \in \mathcal{B}'$ for $i = 1, 2$. Then there exists a linear and base-preserving map T with $T(b_i) = c'_i$ such that the image $T(\mathcal{B})$ is the line segment between c'_1 and c'_2 . One can easily see that $M := \sup(c'_1/c'_2) = (1 - \mu_1)/(1 - \mu_2)$, $m := \inf(c'_1/c'_2) = \mu_1/\mu_2$ and $\Delta(T) = \mathfrak{h}_{\mathcal{C}'}(c'_1, c'_2) = \ln(M/m)$. One can now choose any λ_i with $\lambda_1 \geq \lambda_2 > e^{-\Delta} \lambda_1 > 0$, then set $v := \lambda_1 b_1 - \lambda_2 b_2$, and finally fix μ_i such that $m = e^{-\Delta/2} \sqrt{\lambda_2/\lambda_1}$ and $M = e^{\Delta/2} \sqrt{\lambda_2/\lambda_1}$, which in particular yields $\Delta(T) = \Delta$ and allows one to compute $\mathcal{N}_{\mathcal{B}}(v) = \lambda_2 > 0$ and $\mathcal{N}_{\mathcal{B}'}(T(v)) = \lambda_2 \mu_2 - \lambda_1 \mu_1 > 0$, ensuring $T(v) \notin \mathcal{C}'$. The negativity contraction ratio is then, after some simplification,

$$\frac{\mathcal{N}_{\mathcal{B}'}(T(v))}{\mathcal{N}_{\mathcal{B}}(v)} = \frac{1}{e^{\Delta} - 1} \left(e^{\Delta/2} - \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^2.$$

This indeed equals $\tanh[\Delta/4]$ for the choice $\lambda_1 = \lambda_2$ and so incidentally shows that, besides Proposition 8, also the bound in Corollary 9 is tightest possible. Similarly,

$$\frac{\|T(v)\|_{\mathcal{B}'}}{\|v\|_{\mathcal{B}}} = \tanh[\Delta/2] - \frac{2}{\cosh[\Delta]} \frac{\sqrt{\lambda_1 \lambda_2}}{\lambda_1 + \lambda_2} \left[(e^{\Delta/4} - e^{-\Delta/4})^2 - \left(\left(\frac{\lambda_1}{\lambda_2} \right)^{1/4} - \left(\frac{\lambda_1}{\lambda_2} \right)^{-1/4} \right)^2 \right],$$

showing that (39) is indeed optimal, as for a sequence of choices with $\lambda_1/\lambda_2 \nearrow e^\Delta$ this approaches $\tanh[\Delta/2]$.

By a very similar construction one can see that also the upper bounds in Proposition 7 are tightest possible, if they are to depend solely on Hilbert's projective metric. More indirectly, this optimality can also be seen from the derivation (43), since a tighter upper bound in (21) would lead to a tighter upper bound in (43) and contradict the optimality of Corollary 9 established above.

- ¹ Alfsen, E. M., *Compact Convex sets and Boundary Integrals* (Springer, New York, 1971).
- ² Audenaert, K. M. R., Nussbaum, M., Szkola, A., and Verstraete, F., "Asymptotic error rates in quantum hypothesis testing," *Commun. Math. Phys.* **279**, 251–283 (2008).
- ³ Audenaert, K., Plenio, M. B., and Eisert, J., "Entanglement cost under positive-partial-transpose-preserving operations," *Phys. Rev. Lett.* **90**, 027901 (2003).
- ⁴ Alberti, P. M. and Uhlmann, A., "A problem relating to positive linear maps on matrix algebras," *Rep. Math. Phys.* **18**, 163–176 (1980).
- ⁵ Bauer, F. L., "An elementary proof of the Hopf inequality for positive operators," *Numer. Math.* **7**, 331–337 (1965).
- ⁶ Bennett, C. H., DiVincenzo, D. P., Fuchs, C. A., Mor, T., Rains, E., Shor, P. W., Smolin, J. A., and Wootters, W. K., "Quantum nonlocality without entanglement," *Phys. Rev. A* **59**, 1070–1091 (1999).
- ⁷ Birkhoff, G., "Extensions of Jentzsch's theorem," *Trans. Am. Math. Soc.* **85**, 219–227 (1957).
- ⁸ Bushell, P. J., "Hilbert's metric and positive contraction mappings in a Banach space," *Arch. Ration. Mech. Anal.* **52**, 330–338 (1973).
- ⁹ Bushell, P. J., "On the projective contraction ratio for positive linear mappings," *J. Lond. Math. Soc.* **6**, 256–258 (1973).
- ¹⁰ Boyd, S. and Vandenberghe, L., *Convex Optimization* (Cambridge University Press, Cambridge, England, 2004).
- ¹¹ Chefles, A., Jozsa, R., and Winter, A., "On the existence of physical transformations between sets of quantum states," *Int. J. Quantum Inf.* **2**, 11–21 (2004).
- ¹² Datta, N., "Min- and max-relative entropies and a new entanglement monotone," *IEEE Trans. Inf. Theory* **55**, 2816–2826 (2009).
- ¹³ DiVincenzo, D. P., Leung, D., and Terhal, B. M., "Quantum data hiding," *IEEE Trans. Inf. Theory* **48**, 580–598 (2002).
- ¹⁴ Eveson, S. P., "Hilbert's projective metric and the spectral properties of positive linear operators," *Proc. London Math. Soc.* **70**, 411–440 (1995).
- ¹⁵ Fuchs, C. A., *Distinguishability and accessible information in quantum theory*, Ph.D. dissertation, University of New Mexico (1996); e-print arXiv:quant-ph/9601020.
- ¹⁶ Helstrom, C. W., *Quantum Detection and Estimation Theory* (Academic, New York, 1976).
- ¹⁷ Horodecki, M., Horodecki, P., and Horodecki, R., "Separability of mixed states: necessary and sufficient conditions," *Phys. Lett. A* **223**, 1–8 (1996).
- ¹⁸ Horodecki, M., Horodecki, P., and Horodecki, R., "Mixed-state entanglement and distillation: Is there a "bound" entanglement in nature?," *Phys. Rev. Lett.* **80**, 5239–5242 (1998).
- ¹⁹ Horodecki, R., Horodecki, P., Horodecki, M., and Horodecki, K., "Quantum entanglement," *Rev. Mod. Phys.* **81**, 865–942 (2009).
- ²⁰ Horn, R. A. and Johnson, C. R., *Matrix Analysis* (Cambridge University Press, Cambridge, England, 1985).
- ²¹ Holevo, A. S., "Statistical decision theory for quantum systems," *J. Multivariate Anal.* **3**, 337–394 (1973).
- ²² Hopf, E., "An inequality for positive linear integral operators," *J. Math. Mech.* **12**, 683–692 (1963).
- ²³ Kohlberg, E. and Pratt, J. W., "The contraction mapping approach to the Perron-Frobenius theory: why Hilbert's metric?" *Math. Oper. Res.* **7**, 198–210 (1982).
- ²⁴ Lim, Y., "Hilbert's projective metric on Lorenz cones and Birkhoff formula for Lorentzian compressions," *Numer. Linear Algebra Appl.* **423**, 246–254 (2007).
- ²⁵ Li, C.-K. and Poon, Q.-Y., "Interpolation problems by completely positive maps," e-print arXiv:1012.1675 [math.FA].
- ²⁶ Matthews, W. and Winter, A., "On the Chernoff distance for asymptotic LOCC discrimination of bipartite quantum states," *Commun. Math. Phys.* **285**, 161–174 (2009).
- ²⁷ Matthews, W., Wehner, S., and Winter, A., "Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding," *Commun. Math. Phys.* **291**, 813–843 (2009).
- ²⁸ Nielsen, M. A. and Chuang, I. L., *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).
- ²⁹ Peres, A., "Separability criterion for density matrices," *Phys. Rev. Lett.* **77**, 1413–1415 (1996).
- ³⁰ Plenio, M. B., "Logarithmic negativity: A full entanglement monotone that is not convex," *Phys. Rev. Lett.* **95**, 090503 (2005).
- ³¹ Rains, E. M., "Entanglement purification via separable superoperators," e-print arXiv:quant-ph/9707002.
- ³² Rockafellar, R. T., *Convex Analysis* (Princeton University, Princeton, NJ, 1970).

- ³³Ruskai, M. B., "Beyond strong subadditivity? Improved bounds on the contraction of generalized relative entropy," *Rev. Math. Phys.* **6**, 1147–1161 (1994).
- ³⁴Temme, K., Kastoryano, M. J., Ruskai, M. B., Wolf, M. M., and Verstraete, F., "The χ^2 -divergence and mixing times of quantum Markov processes," *J. Math. Phys.* **51**, 122201 (2010).
- ³⁵Uhlmann, A., "The "transition probability" in the state space of a *-algebra," *Rep. Math. Phys.* **9**, 273–279 (1976).
- ³⁶Vidal, G. and Tarrach, R., "Robustness of entanglement," *Phys. Rev. A* **59**, 141–155 (1999).
- ³⁷Vidal, G. and Werner, R. F., "Computable measure of entanglement," *Phys. Rev. A* **65**, 032314 (2002).
- ³⁸Werner, R. F., "Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model," *Phys. Rev. A* **40**, 4277–4281 (1989).
- ³⁹Note that the terminology appearing in this passage is not entirely unique throughout the literature. In particular, the meaning of *pointed* and *proper* varies from author to author, and *solid* is named in a number of different ways.
- ⁴⁰Also, we will sometimes call \mathfrak{h} 'Hilbert's metric' for short, and $\mathfrak{h}(a, b)$ the 'Hilbert distance' (between a and b).
- ⁴¹In general, $\mathfrak{h}(a, b) < \infty$ holds if and only if both a and b are interior to the intersection of the line through them with \mathcal{C} . That is, the distance between two boundary points can be finite if they are elements of the same exposed face.
- ⁴²The logarithm here is often taken with respect to base 2, e.g., when measuring information in *bits* see also Ref. 37. However, it is necessary to use the natural logarithm in the definition (8) of Hilbert's metric in order for the statements in this paper to hold. Note also the natural logarithm in Eq. (41).

Summary and outlook

7.1 Summaries

Chapter 3: quantum Markov chain mixing

In chapter 3, we have seen that by generalizing the χ^2 -divergence to the quantum setting, many of the classical results for the convergence of Markov processes can be recovered. The general perception, that the convergence should be governed by the spectral properties of the quantum channel could be verified in the asymptotic limit. The fact that we were working with non-commuting probabilities gave rise to a larger set of possibilities of defining an inversion of the fixed point density matrix, all of which lead to a valid upper bound for the trace distance. An interesting question is: how do the different singular values s_i^k of the corresponding quantum discriminant relate to each other? The generalization of the χ^2 -divergence also led to the definition of detailed balance for quantum channels. Again, no single condition for quantum detailed balance exists but an entire family of conditions each determined by a different function $k \in \mathcal{K}$, all of which coincide in the case when we consider classical stochastic processes on a commuting subspace. The quantum concept of detailed balance therefore appears to be richer and allows for a wider set of channels to obey this definition. The conductance bound that was derived could only be shown for unital quantum channels. However we would like to point out, that it is possible to give conductance bounds for classical maps when the Markov chain is not doubly stochastic. The fact that in general we may not assume that the fixed point of an arbitrary channel commutes with the eigenvector associated to the second largest eigenvalues seems to hinder a generalization for non-unital channels. Moreover, the classical conductance bound has a nice geometrical interpretation in terms of the cut set analysis and the maximal flow on the graph associated to the stochastic matrix P_{ij} . When investigating general quantum channels such a nice geometric interpretation seems to be lacking. For unital quantum channels, Cheeger's constant can also be viewed in terms of the minimal probability flow of one subspace to its complement.

Chapter 4: the cutoff phenomenon

In chapter 4, we have introduced the notion of the Cutoff Phenomenon in the context of quantum information theory and applied it to analyze the convergence behavior of some composite quantum processes in continuous time. In particular, we show that the convergence, measured in the trace-norm, of a tensor product of one-parameter semigroups of time evolutions always exhibits cutoff-type behavior. We identify two specific cases (primitive channels with separable initial states, and channels with a unique pure fixed point), which exhibit a true cutoff.

We conclude by noting two directions where the methods introduced in this chapter could be of use. The first is the task of passive error protection in the presence of local noise. It was shown recently [PKSC09] that, if the noise is locally depolarizing and allowing for arbitrary Hamiltonian control, an optimal protection time of order $\mathcal{O}(\log n)$ can be achieved. Theorem 45 gives a strict upper bound on the amount of time that one bit of classical (and hence also quantum) information can be encoded into n qubits, when every qubit is subjected to local noise, and no Hamiltonian control is allowed for. The upper bound happens to coincide with the one in [PKSC09], indicating that their result might not be restricted to depolarizing channels, but could be a general feature of tensor product channels. Along similar lines, a second extension of the above results is in the study of continuous time quantum information theory, where channels are replaced by one-parameter semigroups, and standard objects, such as channel capacities and compression rates, become functions of time.

Chapter 5: dissipative engineering

In chapter 5, we considered three tasks of dissipative engineering. The first proposed a feasible scheme for preparing a maximally entangled state of two atoms in a cavity QED setup. We showed, both analytically and numerically, that the scheme is rapid and reliable, and that the scaling of the fidelity is better than any known closed system protocol. This is a strong indication that thinking of problems from a dissipative engineering perspective can lead to fundamental improvements over protocols which strictly fit within the closed system paradigm. The second task proposes a method for preparing graph states dissipatively, and gives the exact worst case scaling behavior. It is shown that the convergence exhibits a cutoff at time $\mathcal{O}(\log n)$.

The third task we consider is dissipative quantum computation. We adapt the Kitaev clock trick to the dissipative setting in order to engineer a set of Lindblad operators which drive the system into a unique stationary state, in a time which is polynomial in the system size, such that the outcome of the computation can be read out efficiently from the stationary state. We thus show that any circuit quantum computation can be mapped onto dissipative quantum computation with only polynomial overhead. By adapting Kitaev's unary representation of the clock, we are able to construct a 5-local master equation which performs the desired computation, and prove that it converges rapidly.

Chapter 6: Hilbert’s projective metric

In chapter 6, we introduced Hilbert’s projective metric into quantum information theory, where different convex sets and cones appear (such as the cones of positive semidefinite or of separable matrices), and where corresponding cone-preserving maps are ubiquitous (e.g. completely positive maps or LOCC operations). We have found connections and applications to entanglement measures, via base norms and negativities, and to measures for statistical distinguishability of quantum states.

In particular, the projective diameter of a quantum channel yields contraction bounds for distinguishability measures and for entanglement measures under application of the channel. Such non-trivial contraction coefficients are hard to obtain by other means. For instance, whereas the second-largest eigenvalue of a channel determines its asymptotic contraction rates, the same is not true for its finite-time contraction behavior (albeit frequently assumed so). The projective diameter, however, yields valid contraction ratios even for the initial time.

These contraction results may sometimes be tools of more theoretical than practical interest, e.g. by being a guarantee for strict exponential contractivity. This is because, on the one hand, Hilbert’s projective metric $h_{\mathcal{C}}(a, b)$ is efficiently computable given an efficient description of \mathcal{C} . On the other hand, however, the definition of the projective diameter $\Delta(T)$ does not directly entail convex optimization. We have seen examples where $\Delta(T)$ was exactly computable and other examples where this seemed not easy. Nevertheless, even non-trivial upper bounds on $\Delta(T)$ yield non-trivial contraction ratios and ensure immediate exponential convergence.

Besides these contractivity results, Hilbert’s projective metric w.r.t. the positive semidefinite cone decides the possibility of extending a completely positive map, thereby yielding an operational interpretation.

7.2 Outlook

The work in this thesis laid a few preliminary bricks of a theory which is still very young and incomplete. It is reasonable to expect, given its importance in classical sampling problems, that the theory of quantum Markov chain mixing will see some important progress in the near future. What it has especially lacked so far have been explicit examples and explicit bounds on relevant physical processes. This can be attributed on the one hand to the fact that quantum systems are intrinsically complicated, and the types of statements we would like to make are asymptotic, so that numerics cannot guide our intuition very far. On the other hand, trying to generalize classical examples can be misleading, as the classical proofs often rely very heavily on geometric intuitions from graph theory which often do not have any obvious quantum counterparts.

A number of areas are promising for the near future. As a natural generalization of the work in [TKR⁺10], one can consider quantum Log-Sobolev inequalities for finite systems. This would indeed complete the program of extending the basic functional techniques of Markov chain mixing to the quantum setting. With a proper theory of quantum Log-Sobolev inequalities, one could also consider extending the work of Martinelli on

mixing analysis of the classical 2D Ising model to quantum mixing of the transverse (quantum) 2D Ising model.

Applications of mixing time techniques to the study of many-body phenomena is also very promising. Two possible lines of attack are; (i) One can exploit the correspondence between quantum channels and finitely correlated states and apply results from quantum channel mixing to give rigorous bounds on the correlation lengths of many body states. It would be particularly interesting to incorporate notions of symmetry and relate them to correlation lengths and stability of phases. (ii) Another promising direction is to study problems of computational complexity from the open systems perspective. Dissipative quantum computation offers a significantly new approach to computation, analogous to cellular automaton computation in the classical setting, which could lead to new insight for instance in formulating new QMA complete problems.

Finally, the experimental proposals and tests of ideas from dissipative engineering are so far very limited, and there remains a lot of uncharted territory to explore, in terms of state preparation, but also for more sophisticated protocols such as dissipative quantum repeaters or dissipative memories.

Bibliography

- [AD86] David Aldous and Persi Diaconis. Shuffling Cards and Stopping Times. *The American Mathematical Monthly*, 5:333–348, 1986.
- [ADR82] Alain Aspect, Jean Dalibard, and Gérard Roger. Experimental Test of Bell’s Inequalities Using Time- Varying Analyzers. *Physical Review Letters*, 49(25):1804–1807, December 1982.
- [AEM⁺95] M H Anderson, J R Ensher, M R Matthews, C E Wieman, and E A Cornell. Observation of bose-einstein condensation in a dilute atomic vapor. *Science (New York, N.Y.)*, 269(5221):198–201, July 1995.
- [AF] David Aldous and James Allen Fill. Reversible Markov Chains and Random Walks on Graphs.
- [AFH09] R Alicki, M Fannes, and M Horodecki. On thermalization in Kitaev’s 2D model. *Journal of Physics A: Mathematical and Theoretical*, 42(6):065303, February 2009.
- [AGR81] Alain Aspect, Philippe Grangier, and Gérard Roger. Experimental Tests of Realistic Local Theories via Bell’s Theorem. *Physical Review Letters*, 47(7):460–463, August 1981.
- [AGR82] Alain Aspect, Philippe Grangier, and Gérard Roger. Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell’s Inequalities. *Physical Review Letters*, 49(2):91–94, July 1982.
- [AHHH08] R. Alicki, M. Horodecki, P. Horodecki, and R. Horodecki. On thermal stability of topological qubit in Kitaev’s 4D model. *arXiv:0811.0033*, November 2008.
- [Alb83] Peter M. Alberti. A note on the transition probability over C*-algebras. *Letters in Mathematical Physics*, 7(1):25–32, January 1983.
- [Ali76] R Alicki. On the detailed balance condition for non-hamiltonian systems. *Reports on Mathematical Physics*, 10(2):249–258, October 1976.
- [AN00] A. Amari and H. Nagaoka. *Methods of Information Geometry*. American Mathematical Society and Oxford University Press, 2000.

- [ANSV08] K. M. R. Audenaert, M. Nussbaum, A. Szkola, and F. Verstraete. Asymptotic Error Rates in Quantum Hypothesis Testing. *Communications in Mathematical Physics*, 279(1):251–283, February 2008.
- [Ara76] Huzihiro Araki. Relative entropy of state of von Neumann algebras. *Publ. RIMS Kyoto Univ.*, 9:809–833, 1976.
- [Ara87] Huzihiro Araki. Recent progress in entropy and relative entropy. In R. Seaneor and M. Mebkhout, editors, *VIIIth International Congress on Mathematical Physics*, pages 354–365. World Scientific, Singapore, 1987.
- [BD92] Dave Bayer and Persi Diaconis. Trailing the Dovetail Shuffle to its Lair. *Annals of Applied Probability*, 2:294–313, 1992.
- [BDI⁺11] J. Busch, S. De, S. Ivanov, B. Torosov, T. Spiller, and A. Beige. Cooling atom-cavity systems into entangled states. *Physical Review A*, 84(2), August 2011.
- [Bek66] John Beka. On the Problem of Hidden Variables in Quantum Mechanics. *Reviews of Modern Physics*, 38(3):447–452, July 1966.
- [BEO09] Christian Burrell, Jens Eisert, and Tobias Osborne. Information propagation through quantum chains with fluctuating disorder. *Physical Review A*, 80(5), November 2009.
- [BH11] Sergey Bravyi and Jeongwan Haah. Energy Landscape of 3D Spin Hamiltonians with Topological Order. *Physical Review Letters*, 107(15), October 2011.
- [Bha97] R. Bhatia. *Matrix Analysis*. Springer-Verlag, New York, 1997.
- [BKNPV08] Robin Blume-Kohout, Hui Ng, David Poulin, and Lorenza Viola. Characterizing the Structure of Preserved Information in Quantum Processes. *Physical Review Letters*, 100(3), January 2008.
- [BLY06] J Barrera, B Lachaud, and B Ycart. Cut-off for nn-tuples of exponentially converging processes. *Stochastic Processes and their Applications*, 116(10):1433–1446, October 2006.
- [BMS⁺11] Julio T Barreiro, Markus Müller, Philipp Schindler, Daniel Nigg, Thomas Monz, Michael Chwalla, Markus Hennrich, Christian F Roos, Peter Zoller, and Rainer Blatt. An open-system quantum simulator with trapped ions. *Nature*, 470(7335):486–91, March 2011.
- [BO07] Christian Burrell and Tobias Osborne. Bounds on the Speed of Information Propagation in Disordered Quantum Spin Chains. *Physical Review Letters*, 99(16), October 2007.
- [Boh35] N. Bohr. Can Quantum-Mechanical Description of Physical Reality be Considered Complete? *Physical Review*, 48(8):696–702, October 1935.
- [Bro11] L. E. J. Brouwer. Über Abbildung von Mannigfaltigkeiten. *Mathematische Annalen*, 71(1):97–115, March 1911.

- [CL93] Eric A. Carlen and Elliott H. Lieb. Optimal hypercontractivity for fermi fields and related non-commutative integration inequalities. *Communications in Mathematical Physics*, 155(1):27–46, July 1993.
- [CLBT10] Stefano Chesi, Daniel Loss, Sergey Bravyi, and Barbara M Terhal. Thermodynamic stability criteria for a quantum memory based on stabilizer and subsystem codes. *New Journal of Physics*, 12(2):025013, February 2010.
- [CV09] J Ignacio Cirac and Frank Verstraete. Renormalization and tensor product states in spin chains and lattices. *Journal of Physics A: Mathematical and Theoretical*, 42(50):504004, December 2009.
- [CW98] Eric A. Cornell and Carl E. Wieman. The Bose-Einstein Condensate. *Scientific American*, 278(3):40–45, March 1998.
- [CZ95] J. Cirac and P. Zoller. Quantum Computations with Cold Trapped Ions. *Physical Review Letters*, 74(20):4091–4094, May 1995.
- [Dat09] Nilanjana Datta. Min- and Max-Relative Entropies and a New Entanglement Monotone. *IEEE Transactions on Information Theory*, 55(6):2816–2826, June 2009.
- [DFK91] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM*, 38(1):1–17, January 1991.
- [Dia96] Persi Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Natl. Acad. Sci. U.S.A.*, 93:1659, 1996.
- [Dia98] P Diaconis. What Do We Know about the Metropolis Algorithm ?,. *Journal of Computer and System Sciences*, 57(1):20–36, August 1998.
- [Dia11] Persi Diaconis. The Mathematics of Mixing Things Up. *Journal of Statistical Physics*, August 2011.
- [DLP08] Jian Ding, Eyal Lubetzky, and Yuval Peres. Total variation cutoff in birth-and-death chains. *Probability Theory and Related Fields*, 146(1-2):61–85, November 2008.
- [DMA⁺95] K. Davis, M. Mewes, M. Andrews, N. van Druten, D. Durfee, D. Kurn, and W. Ketterle. Bose-Einstein Condensation in a Gas of Sodium Atoms. *Physical Review Letters*, 75(22):3969–3973, November 1995.
- [DMK⁺08] S. Diehl, A. Micheli, A. Kantian, B. Kraus, H. P. Büchler, and P. Zoller. Quantum states and phases in driven open quantum systems with cold atoms. *Nature Physics*, 4(11):878–883, September 2008.
- [DRBH95] P. Domokos, J. Raimond, M. Brune, and S. Haroche. Simple cavity-QED two-bit universal quantum logic gate: The principle and expected performances. *Physical Review A*, 52(5):3554–3559, November 1995.
- [DRBZ11] Sebastian Diehl, Enrique Rico, Mikhail A. Baranov, and Peter Zoller. Topology by dissipation in atomic quantum wires. *Nature Physics*, October 2011.

- [DS91] Persi Diaconis and Daniel Stroock. Geometric Bounds for Eigenvalues of Markov Chains. *The Annals of Applied Probability*, 1(1):36–61, February 1991.
- [DSC93] Persi Diaconis and Laurent Saloff-Coste. Comparison Theorems for Reversible Markov Chains. *The Annals of Applied Probability*, 3(3):696–730, August 1993.
- [DTM⁺10] Sebastian Diehl, Andrea Tomadin, Andrea Micheli, Rosario Fazio, and Peter Zoller. Dynamical Phase Transitions and Instabilities in Open Atomic Many-Body Systems. *Physical Review Letters*, 105(1), July 2010.
- [EHK78] D. E. Evans and R. Hoegh-Krohn. Spectral Properties of Positive Maps on C*-Algebras. *Journal of the London Mathematical Society*, s2-17(2):345–355, April 1978.
- [EPR35] A. Einstein, B. Podolsky, and N. Rosen. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Physical Review*, 47(10):777–780, May 1935.
- [Fey82] Richard P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6-7):467–488, June 1982.
- [FH91] K. H. Fischer and J. A. Hertz. *Spin Glasses*. Cambridge University Press, 1991.
- [Fil91] James Allen Fill. Eigenvalue Bounds on Convergence to Stationarity for Nonreversible Markov Chains, with an Application to the Exclusion Process. *The Annals of Applied Probability*, 1(1):62–87, February 1991.
- [FNW92] M. Fannes, B. Nachtergaele, and R. F. Werner. Finitely correlated states on quantum spin chains. *Communications in Mathematical Physics*, 144(3):443–490, March 1992.
- [Fri78] Alberto Frigerio. Stationary states of quantum dynamical semigroups. *Communications in Mathematical Physics*, 63(3):269–276, October 1978.
- [Fri90] Alberto Frigerio. Simulated annealing and quantum detailed balance. *Journal of Statistical Physics*, 58(1-2):325–354, January 1990.
- [Fro12] Georg Frobenius. Ueber Matrizen aus nicht negativen Elementen. *Sitzungsber. Königl. Preuss. Akad. Wiss.*, pages 456–477, 1912.
- [Gac86] P Gacs. Reliable computation with cellular automata. *Journal of Computer and System Sciences*, 32(1):15–78, February 1986.
- [Gac01] Peter Gacs. Reliable Cellular Automata with Self-Organization. *Journal of Statistical Physics*, 103(1-2):45–267, 2001.
- [Has07] M. Hastings. Random unitaries give quantum expanders. *Physical Review A*, 76(3), September 2007.

- [HDE⁺05] M. Hein, W. Dür, J. Eisert, R. Raussendorf, M. Van den Nest, and H. J. Briegel. Entanglement in Graph States and its Applications. *Proceedings of the International School of Physics "Enrico Fermi" on "Quantum Computers, Algorithms and Chaos"*, Varenna, Italy, July, 2005, page 99, February 2005.
- [Hia93] F Hiai. The Golden-Thompson trace inequality is complemented. *Linear Algebra and its Applications*, 181:153–185, March 1993.
- [HJ06] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 2006.
- [HJ07] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2007.
- [HN99] Olle Haggstrom and Karin Nelander. On Exact Simulation of Markov Random Fields Using Coupling from the Past. *Scandinavian Journal of Statistics*, 26(3):395–411, September 1999.
- [HP91] Fumio Hiai and Denes Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Communications in Mathematical Physics*, 143(1):99–114, December 1991.
- [Jer03] Mark Jerrum. *Counting, sampling and integrating: algorithms and complexity*. Birkhauser Verlag, Basel, 2003.
- [JS88] Mark Jerrum and Alistair Sinclair. Conductance and the rapid mixing property for Markov chains: the approximation of permanent resolved. In *Proceedings of the twentieth annual ACM symposium on Theory of computing - STOC '88*, pages 235–244, New York, New York, USA, 1988. ACM Press.
- [JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM*, 51(4):671–697, July 2004.
- [KBD⁺08] B. Kraus, H. Büchler, S. Diehl, A. Kantian, A. Micheli, and P. Zoller. Preparation of entangled states by quantum Markov processes. *Physical Review A*, 78(4), October 2008.
- [KMJ⁺11] Hanna Krauter, Christine Muschik, Kasper Jensen, Wojciech Wasilewski, Jonas Petersen, J. Cirac, and Eugene Polzik. Entanglement Generated by Dissipation and Steady State Entanglement of Two Macroscopic Objects. *Physical Review Letters*, 107(8), August 2011.
- [KRSr11] M. Kastoryano, F. Reiter, and A. S. Sørensen. Dissipative Preparation of Entanglement in Optical Cavities. *Physical Review Letters*, 106(9), February 2011.
- [KRW11] Michael J. Kastoryano, David Reeb, and Michael M. Wolf. A Cutoff Phenomenon for Quantum Markov Chains. *arXiv:1111.2123*, November 2011.
- [KSV00] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi. *Classical and Quantum Computation*. Graduate Studies in Mathematics, Vol. 47, AMS, 2000.

- [LBMW03] D. Leibfried, R. Blatt, C. Monroe, and D. Wineland. Quantum dynamics of single trapped ions. *Reviews of Modern Physics*, 75(1):281–324, March 2003.
- [LCW⁺96] C. Livermore, C. H. Crouch, R. M. Westervelt, K. L. Campman, and A. C. Gossard. The Coulomb Blockade in Coupled Quantum Dots. *Science*, 274(5291):1332–1335, November 1996.
- [LD98] Daniel Loss and David P. DiVincenzo. Quantum computation with quantum dots. *Physical Review A*, 57(1):120–126, January 1998.
- [LGL⁺11] Peng-Bo Li, Shao-Yan Gao, Hong-Rong Li, Sheng-Li Ma, and Fu-Li Li. Dissipative preparation of entangled states between two spatially separated nitrogen-vacancy centers. *arXiv:1110.6718*, October 2011.
- [LHN⁺11] B. P. Lanyon, C. Hempel, D. Nigg, M. Muller, R. Gerritsma, F. Zahringer, P. Schindler, J. T. Barreiro, M. Rambach, G. Kirchmair, M. Hennrich, P. Zoller, R. Blatt, and C. F. Roos. Universal Digital Quantum Simulation with Trapped Ions. *Science*, September 2011.
- [Lie73] E Lieb. Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Advances in Mathematics*, 11(3):267–288, December 1973.
- [Lin99] G. Lindblad. A General No-Cloning Theorem. *Letters in Mathematical Physics*, 47(2):189–196, 1999.
- [LPW08] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI, 2008.
- [LR99] Andrew Lesniewski and Mary Beth Ruskai. Monotone Riemannian metrics and relative entropy on noncommutative probability spaces. *Journal of Mathematical Physics*, 40(11):5702, 1999.
- [LS09] Eyal Lubetzky and Allan Sly. Cutoff for the Ising model on the lattice. *arXiv:0909.4320*, page 34, September 2009.
- [Maj84] W. A. Majewski. The detailed balance condition in quantum statistical mechanics. *Journal of Mathematical Physics*, 25(3):614, 1984.
- [Mar94] F. Martinelli. On the two-dimensional dynamical Ising model in the phase coexistence region. *Journal of Statistical Physics*, 76(5-6):1179–1246, September 1994.
- [MC89] E.A. Morozova and N.N. Chentsov. Markov invariant geometry on state manifolds. *Itogi Nauki i Tekhniki. Ser. Sovrem. Probl. Mat. Nov. Dostizh.*, 36:69–102, 1989.
- [Mer04] N. David Mermin. Could Feynman Have Said This? *Physics Today*, page 10, 2004.
- [Mih89] M. Mihail. *No Title*. PhD thesis, Harvard University, 1989.

- [MMK⁺95] C. Monroe, D. Meekhof, B. King, W. Itano, and D. Wineland. Demonstration of a Fundamental Quantum Logic Gate. *Physical Review Letters*, 75(25):4714–4717, December 1995.
- [MO94] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. *Communications in Mathematical Physics*, 161(3):447–486, April 1994.
- [MO08] Ashley Montanaro and Tobias J. Osborne. Quantum boolean functions. *arXiv:0810.2435*, October 2008.
- [MOS94] F. Martinelli, E. Olivieri, and R. H. Schonmann. For 2-D lattice spin systems weak mixing implies strong mixing. *Communications in Mathematical Physics*, 165(1):33–47, October 1994.
- [MPC11] Christine Muschik, Eugene Polzik, and J. Cirac. Dissipatively driven entanglement of two macroscopic atomic ensembles. *Physical Review A*, 83(5), May 2011.
- [MS98] W A Majewski and R F Streater. Detailed balance and quantum dynamical maps. *Journal of Physics A: Mathematical and General*, 31(39):7981–7995, October 1998.
- [NC00] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2000.
- [NO00] H. Nagaoka and T. Ogawa. Strong converse and Stein’s lemma in quantum hypothesis testing. *IEEE Transactions on Information Theory*, 46(7):2428–2433, 2000.
- [OP04] Masanori Ohya and Dénes Petz. *Quantum entropy and its use*. Springer, 2004.
- [Osb11] Tobias J. Osborne. Hamiltonian complexity. *arXiv:1106.5875*, June 2011.
- [OZ99] R Olkiewicz and B Zegarlinski. Hypercontractivity in Noncommutative Lp-Spaces. *Journal of Functional Analysis*, 161(1):246–285, January 1999.
- [Pag65] V . I. Pagmova. An asymptotic formula for the incomplete gamma function. *Engl. Transl. in U.S.S.R. Comput. Math. and Math. Phys.*, 5:162–166, 1965.
- [Pau03] V. Paulsen. *Completely bounded maps and operator spaces*. Cambridge University Press, 2003.
- [PCC11] Fernando Pastawski, Lucas Clemente, and Juan Cirac. Quantum memories based on engineered dissipation. *Physical Review A*, 83(1), January 2011.
- [PCZ96] J. Poyatos, J. Cirac, and P. Zoller. Quantum Reservoir Engineering with Laser Cooled Trapped Ions. *Physical Review Letters*, 77(23):4728–4731, December 1996.

- [Pea00] K. Pearson. On the Criterion that a given system of deviations from the probable in the case of correlated system of variables is such that it can be reasonable supposed to have arisen from random sampling. *Philosophical Magazine*, 50:157–172, 1900.
- [Per07] Oskar Perron. Zur Theorie der Matrices. *Mathematische Annalen*, 64(2):248–263, June 1907.
- [Pet86] D Petz. Quasi-entropies for finite quantum systems. *Reports on Mathematical Physics*, 23(1):57–65, February 1986.
- [Pet96] Denes Petz. Monotone metrics on matrix spaces. *Linear Algebra and its Applications*, 244:81–96, September 1996.
- [PG10] Denes Petz and Catalin Ghinea. Introduction to quantum Fisher information. *arXiv:1008.2417*, August 2010.
- [PGVWC07] David Perez-Garcia, Frank Verstraete, Michael M. Wolf, and Juan I. Cirac. Matrix Product State Representations. *Quantum Inf. Comput.*, 7:401, 2007.
- [PGWPR06] David Perez-Garcia, Michael M. Wolf, Denes Petz, and Mary Beth Ruskai. Contractivity of positive and trace-preserving maps under L_p norms. *Journal of Mathematical Physics*, 47(8):083506, 2006.
- [PKSC09] Fernando Pastawski, Alastair Kay, Norbert Schuch, and Ignacio Cirac. How Long Can a Quantum Memory Withstand Depolarizing Noise? *Physical Review Letters*, 103(8), August 2009.
- [PR98] Denes Petz and Mary Beth Ruskai. Contraction of Generalized Relative Entropy Under Stochastic Mappings on Matrices. *Infinite Dimensional Analysis, Quantum Probability and Related Topics (IDAQP)*, 1:83–89, 1998.
- [Pro11] Tomaž Prosen. Open XXZ Spin Chain: Nonequilibrium Steady State and a Strict Bound on Ballistic Transport. *Physical Review Letters*, 106(21), May 2011.
- [PS96] Denes Petz and Csaba Sudar. Geometries of quantum states. *Journal of Mathematical Physics*, 37(6):2662, 1996.
- [PW96] James Gary Propp and David Bruce Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9(1-2):223–252, August 1996.
- [PW97] James Gary Propp and David Bruce Wilson. Coupling from the Past: a User’s Guide. In *Microsurveys in discrete probability*, page 181. The American Mathematical Society, 1997.
- [PW09] David Poulin and Pawel Wocjan. Sampling from the Thermal Quantum Gibbs State and Evaluating Partition Functions with a Quantum Computer. *Physical Review Letters*, 103(22), November 2009.
- [RGE11] Arnau Riera, Christian Gogolin, and Jens Eisert. Thermalization in nature and on a quantum computer. *arXiv:1102.2389*, February 2011.

- [RKSr11] Florentin Reiter, Michael J. Kastoryano, and Anders S. Sørensen. Entangled steady-states of two atoms in an optical cavity by engineered decay. *arXiv:1110.1024*, October 2011.
- [RS90] M B Ruskai and F H Stillinger. Convexity inequalities for estimating free energy and relative entropy. *Journal of Physics A: Mathematical and General*, 23(12):2421–2437, June 1990.
- [RT00] Dana Randall and Prasad Tetali. Analyzing Glauber dynamics by comparison of Markov chains. *Journal of Mathematical Physics*, 41(3):1598, 2000.
- [Rus94] Mary Beth Ruskai. Beyond Strong Subadditivity? Improved Bounds on the Contraction of Generalized Relative Entropy. *Reviews of Mathematical Physics*, 6:1147–1161, 1994.
- [RW96] S. Richter and R. F. Werner. Ergodicity of quantum cellular automata. *Journal of Statistical Physics*, 82(3-4):963–998, February 1996.
- [SB08] E. Schrödinger and M. Born. Discussion of Probability Relations between Separated Systems. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31(04):555, October 2008.
- [SC04] Laurent Saloff-Coste. Random Walks on Finite Groups. In A.-S. Sznitman and S.R.S. Varadhan, editors, *Probability on discrete structures*, page 263. Springer Lecture Notes in Mathematical Sciences Volume 110, 2004.
- [Sha01] C. E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3, January 2001.
- [Sho94] P.W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134. IEEE Comput. Soc. Press, 1994.
- [SP10] Thomas H. Seligman and Tomaz Prosen. Third quantization. *arXiv:1011.0625*, November 2010.
- [SPGWC10] Mikel Sanz, David Perez-Garcia, Michael M. Wolf, and Juan I. Cirac. A Quantum Version of Wielandt’s Inequality. *IEEE Transactions on Information Theory*, 56(9):4668–4673, September 2010.
- [Spo77] Herbert Spohn. An algebraic condition for the approach to equilibrium of an open N-level system. *Letters in Mathematical Physics*, 2(1):33–38, August 1977.
- [TD00] Barbara Terhal and David DiVincenzo. Problem of equilibration and the computation of correlation functions on a quantum computer. *Physical Review A*, 61(2), January 2000.
- [TDZ11] Andrea Tomadin, Sebastian Diehl, and Peter Zoller. Nonequilibrium phase diagram of a driven and dissipative many-body system. *Physical Review A*, 83(1), January 2011.

- [THL⁺95] Q. Turchette, C. Hood, W. Lange, H. Mabuchi, and H. Kimble. Measurement of Conditional Phase Shifts for Quantum Logic. *Physical Review Letters*, 75(25):4710–4713, December 1995.
- [TKR⁺10] K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf, and F. Verstraete. The χ^2 -divergence and mixing times of quantum Markov processes. *Journal of Mathematical Physics*, 51(12):122201, 2010.
- [TOV⁺11] K Temme, T J Osborne, K G Vollbrecht, D Poulin, and F Verstraete. Quantum Metropolis sampling. *Nature*, 471(7336):87–90, March 2011.
- [Tur37] A. M. Turing. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1):230–265, January 1937.
- [Uhl76] A Uhlmann. The "transition probability" in the state space of a $*$ -algebra. *Reports on Mathematical Physics*, 9(2):273–279, April 1976.
- [VC04] F. Verstraete and J. I. Cirac. Renormalization algorithms for Quantum-Many Body Systems in two and higher dimensions. *arXiv:cond-mat/0407066*, July 2004.
- [Vid07] G. Vidal. Entanglement Renormalization. *Physical Review Letters*, 99(22), November 2007.
- [VMC08] F. Verstraete, V. Murg, and J.I. Cirac. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Advances in Physics*, 57(2):143–224, March 2008.
- [VMC11] Karl Vollbrecht, Christine Muschik, and J. Cirac. Entanglement Distillation by Dissipation and Continuous Quantum Repeaters. *Physical Review Letters*, 107(12), September 2011.
- [Von45] John Von Neumann. First Draft of a Report on the EDVAC. Technical report, United States Army Ordnance Department, 1945.
- [VWC09] Frank Verstraete, Michael M. Wolf, and J. Ignacio Cirac. Quantum computation and quantum-state engineering driven by dissipation. *Nature Physics*, 5(9):633–636, July 2009.
- [Wat04] John Watrous. Notes on super-operator norms induced by Schatten norms. *arXiv:quant-ph/0411077*, November 2004.
- [Wat08] John Watrous. Quantum Computational Complexity. *arXiv:0804.3401*, April 2008.
- [Wat09] John Watrous. Semidefinite Programs for Completely Bounded Norms. *Theory of Computing*, 5:217–238, 2009.
- [WBB⁺92] D. Wineland, J.C. Bergquist, J.J. Bollinger, W. Itano, L. Moore, J.M. Gilligen, M.G. Raizen, D.J. Heinzen, C.S. Weimer, and C.H. Manney. Recent

- experiments on trapped ions at the National Institute of Standards and Technology. In E. Arimondo, W.D. Phillips, and F. Strumia, editors, *Laser Manipulation of Atoms and Ions*, page 553. Proceedings of the International School of Physics "Enrico Fermi" Course 118,, 1992.
- [WBM⁺95] F. Waugh, M. Berry, D. Mar, R. Westervelt, K. Campman, and A. Gosard. Single-Electron Charging in Double and Triple Quantum Dots with Tunable Coupling. *Physical Review Letters*, 75(4):705–708, July 1995.
- [Wer89] Reinhard Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Physical Review A*, 40(8):4277–4281, October 1989.
- [Wil12] C. O. Williams. A card reading. *The Magician Monthly*, 8:67, 1912.
- [WML⁺10] Hendrik Weimer, Markus Müller, Igor Lesanovsky, Peter Zoller, and Hans Peter Büchler. A Rydberg quantum simulator. *Nature Physics*, 6(5):382–388, March 2010.
- [Wol10] Michael M. Wolf. *Quantum Channels and Operations*. www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf, 2010.
- [WPG10] Michael M. Wolf and David Perez-Garcia. The inverse eigenvalue problem for quantum channels. *arXiv:1005.4545*, May 2010.
- [WS10] Xiaoting Wang and Sophie G. Schirmer. Generating maximal entanglement between non-interacting atoms by collective decay and symmetry breaking. *arXiv:1005.2114*, page 4, May 2010.